

OM of: US-09-303-518d-127 to: SwissProt_40.* out_format: pfs
Date: Jun 30, 2002 8:33 AM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame-azp.model -DEV=xlh
-Q/cgn2.1/USPPO.spool/US09303518/runat_28062002.147714.4362/app.query.fasta.1.23501
-DB=SwissProt_40 -OPMT=fastan -SUFFIX=isp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=dlsum62
-TRAN=human40.cdi -LIST=100 -DOCALLIGN=200 -THR_SCORE=pt
-THR_MAX=100 -THR_MIN=0 -ALIGN=5 -MODE=LOCAL -OUTPRF=pfs
-NORMEXT -HARSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09303518 -ECGN1_1.440 -NCPD=6 -ICPD=3 -LONCLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -NO_XLPYX -WAIT -THREADS=1

Search information block:

Query: US-09-303-518d-127
Database: SwissProt_40.*
Database sequences: 105224
Database length: 38719550
Search time (sec): 217.960000

Score_List:

Sequence	Strd Orig	ZScore	Escore Len	Documentation
SwissProt_40:NORA_NEIMA + 2211.00 3356.88	2.3e-174	447	1	Q91y08 neisseria meningitidis
SwissProt_40:NORA_NEIMA + 2177.00 3306.60	1.5e-171	447	1	Q9K0M1 neisseria meningitidis
SwissProt_40:NORA_PASIM + 1614.50 2374.79	3.1e-125	446	1	Q9C1B1 pasteurella multocida
SwissProt_40:NORA_HAEIN + 1574.00 2314.87	6.8e-122	447	1	P43955 haemophilus influenza
SwissProt_40:NORA_VIBAL + 1424.50 2090.81	1.4e-109	446	1	Q9JFW1 vibrio Harveyi. na(+)
SwissProt_40:NORA_VIBCH + 1412.50 2076.06	2.1e-108	446	1	Q9K2P1 vibrio cholerae. na(+)
SwissProt_40:NORA_PSEAE + 1298.00 1906.76	3.7e-99	445	1	Q9K2P1 pseudomonas aeruginos
SwissProt_40:NORA_CHLTP + 447.50 639.57	1.2e-29	467	1	Q9Z7G2 chlamydia pneumoniae
SwissProt_40:NORA_CHLTR + 445.00 615.34	3.0e-27	465	1	Q9Z7G2 chlamydia muridarum
SwissProt_40:NORA_VIBCH + 157.00 214.17	4.0e-05	774	1	Q9K2P1 vibrio cholerae. elec
SwissProt_40:HKRI_YEAST + 136.00 175.08	0.0026	1802	1	P41809 saccharomyces cerevi
SwissProt_40:NI21_RAT + 123.50 160.47	0.0255	1199	1	P53591 rattus norvegicus (rat
SwissProt_40:AMVH_YEAST + 123.00 158.48	0.0288	1367	1	P08640 saccharomyces cerevi
SwissProt_40:RNCC_PSEAE + 121.00 154.93	0.0372	774	1	Q9HY88 pseudomonas aeruginos
SwissProt_40:VGLX_HSVB + 116.50 150.00	0.0878	797	1	P28968 equine herpesvirus ty
SwissProt_40:RNCC_HAEIN + 116.50 153.74	0.0884	819	1	P71397 haemophilus influenza
SwissProt_40:YN66_YEAST + 110.00 141.18	0.3245	1117	1	P53753 saccharomyces cerevi
SwissProt_40:RNCC_HAEIN + 109.50 148.94	0.2933	457	1	P45061 haemophilus influenza
SwissProt_40:YGC3_HAISO + 108.00 147.73	0.3016	539	1	Q32716 rhododactyl capsulatu
SwissProt_40:AM12_YEAST + 107.50 147.83	0.4106	376	1	P21561 haloterrax sp. (strain
SwissProt_40:AGAL_YEAST + 107.50 141.59	0.4742	725	1	P32323 saccharomyces cerevi
SwissProt_40:NI21_HUMAN + 107.00 135.83	0.5854	1229	1	P09723 homo sapiens (human)
SwissProt_40:ATPB_PYLTI + 106.50 144.01	0.5239	481	1	P23653 saccharomyces cerevi
SwissProt_40:FI22_YEAST + 105.50 131.05	0.8254	1609	1	P23653 saccharomyces cerevi
SwissProt_40:AMVH_PIG + 105.00 133.51	0.8430	1150	1	P12021 sus scrofa (pig). apom
SwissProt_40:PA51_SCHPO + 103.50 125.69	1.128	2073	1	P09020 sus scrofa (pig). apom
SwissProt_40:GID_BACSU + 103.00 139.79	0.9953	435	1	P39815 bacillus subtilis. prc
SwissProt_40:GID_BACSU + 103.00 138.84	1.102	481	1	P30158 bacillus subtilis. prc
SwissProt_40:RNCC_ECOS7 + 103.00 134.74	1.112	740	1	P58383 escherichia coli O157
SwissProt_40:RNCC_ECOS7 + 103.00 134.74	1.112	740	1	P76111 escherichia coli. ele
SwissProt_40:RNCC_ECOS7 + 101.00 132.64	1.160	676	1	P60528 mesocricetus auratus
SwissProt_40:RNCC_ECOS7 + 101.00 127.50	1.80	1161	1	P47179 saccharomyces cerevi
SwissProt_40:RNCC_ECOS7 + 100.50 136.64	1.58	411	1	P87049 schizosaccharomyces f
SwissProt_40:RNCC_ECOS7 + 100.50 135.30	1.63	473	1	P57215 buchnera aphidicola
SwissProt_40:RNCC_ECOS7 + 100.50 121.43	2.24	2035	1	P51610 homo sapiens (human)
SwissProt_40:RNCC_ECOS7 + 100.00 135.29	1.76	438	1	Q949110 mycoplasma capricolun

SwissProt_40:PMAL_DUNAC + 100.00 126.51	2.16	1103	P54210 dunaliella acidop
SwissProt_40:YGLF_YEAST + 99.50 132.37	2.04	551	P53214 saccharomyces cere
SwissProt_40:HFOL_MESAU + 99.00 118.96	3.00	2090	P51611 mesocricetus aura
SwissProt_40:FSHR_MOUSE + 98.50 127.39	2.67	796	P09625 caenorhabditis ele
SwissProt_40:YGBA_CAEBL + 98.00 121.95	3.27	1306	P33234 saccharomyces cer
SwissProt_40:SON_HUMAN + 98.00 116.06	3.75	2426	P18583 homo sapiens (huma
SwissProt_40:YH08_YEAST + 97.50 126.52	3.04	605	P38739 saccharomyces cere
SwissProt_40:CHPV_MYCTU + 97.50 126.23	3.20	770	P76111 escherichia coli. ele
SwissProt_40:HYPF_AZOVI + 97.00 122.98	3.73	1003	P40596 azotobacter vinela
SwissProt_40:SPCC_CHICK + 97.00 122.98	3.73	1003	P12872 g trifunctional p
SwissProt_40:SPCC_CHICK + 96.50 131.27	3.33	1238	P68032 rhododactyl capsu
SwissProt_40:YH08_HUMAN + 96.00 122.92	4.36	864	Q9ULK2 homo sapiens (huma
SwissProt_40:YH08_HUMAN + 95.50 126.37	4.36	864	Q9ULK2 homo sapiens (huma
SwissProt_40:YH08_HUMAN + 95.00 122.15	5.19	802	P29352 mus musculus (mous
SwissProt_40:VATP_ARCPU + 94.50 124.47	5.32	581	Q29101 archaoglobus fulg
SwissProt_40:MCL_MOUSE + 94.00 123.70	5.41	630	Q02496 mus musculus (mous
SwissProt_40:FSHR_MACPA + 94.00 122.03	6.08	695	P32212 macaca fasciculari
SwissProt_40:YH24_CAEBL + 93.50 124.01	6.28	522	Q27245 caenorhabditis ele
SwissProt_40:YH24_CAEBL + 93.00 123.49	6.87	510	P07001 escherichia coli.
SwissProt_40:YH24_CAEBL + 93.00 119.61	7.51	767	P04065 saccharomyces dias
SwissProt_40:YH24_CAEBL + 93.00 112.69	8.81	819	Q9Z7H7 pyrococcus abyssl.
SwissProt_40:YH24_CAEBL + 92.50 115.99	8.05	1589	P39769 pyrococcus abyssl.
SwissProt_40:YH24_CAEBL + 92.50 115.06	9.02	682	P50478 gallus gallus (chi
SwissProt_40:YH24_CAEBL + 92.50 114.15	9.21	1260	P24384 saccharomyces cer
SwissProt_40:YH24_CAEBL + 92.00 123.48	8.03	437	Q9ULV0 homo sapiens (hum
SwissProt_40:YH24_CAEBL + 92.00 122.42	8.04	440	P66372 rhodococcus erythr
SwissProt_40:YH24_CAEBL + 92.00 122.07	8.29	440	O66119 zymomonas mobilis
SwissProt_40:YH24_CAEBL + 92.00 119.04	8.89	507	O53743 mycobacterium tube
SwissProt_40:YH24_CAEBL + 92.00 114.27	9.93	697	P71886 mycobacterium tube
SwissProt_40:YH24_CAEBL + 92.00 114.00	9.99	1152	P27816 homo sapiens (hum
SwissProt_40:YH24_CAEBL + 91.50 119.73	9.46	1185	P23226 dirosophila melano
SwissProt_40:YH24_CAEBL + 91.50 114.88	10.58	600	P14328 dictyostelium disc
SwissProt_40:YH24_CAEBL + 91.00 115.28	11.33	886	P18596 rattus norvegicus (r
SwissProt_40:YH24_CAEBL + 91.00 115.06	11.39	907	P03200 epstein-barr virus
SwissProt_40:YH24_CAEBL + 90.50 115.90	12.07	768	P29760 saccharomyces dias
SwissProt_40:YH24_CAEBL + 90.00 122.16	11.62	368	Q9ULY3 pyrococcus abyssl.
SwissProt_40:YH24_CAEBL + 90.00 122.92	11.99	419	P28909 escherichia coli.
SwissProt_40:YH24_CAEBL + 90.00 119.89	11.90	467	P50770 human papillomaviru
SwissProt_40:YH24_CAEBL + 90.00 116.87	12.76	642	P33299 saccharomyces cere
SwissProt_40:YH24_CAEBL + 89.50 114.58	13.45	817	P42850 pyrococcus furiosu
SwissProt_40:YH24_CAEBL + 89.50 121.68	12.34	358	P39912 b. aroa(g) protein
SwissProt_40:YH24_CAEBL + 89.50 119.17	13.08	466	P54208 synecococcus sp.
SwissProt_40:YH24_CAEBL + 89.50 117.43	13.62	466	P50801 human papillomaviru
SwissProt_40:YH24_CAEBL + 89.50 117.43	13.62	560	P00589 corynebacter beta.
SwissProt_40:YH24_CAEBL + 89.50 116.84	13.80	596	P53379 saccharomyces cere
SwissProt_40:YH24_CAEBL + 89.50 116.84	13.80	728	O9Z512 streptomyces coeli
SwissProt_40:YH24_CAEBL + 89.00 117.80	14.59	498	P26529 nicotiana glumigil
SwissProt_40:YH24_CAEBL + 89.00 117.80	14.59	498	P26530 nicotiana glumigil
SwissProt_40:YH24_CAEBL + 89.00 116.88	14.90	549	P15337 escherichia coli.
SwissProt_40:YH24_CAEBL + 88.50 112.96	15.81	829	P44323 haemophilus influen
SwissProt_40:YH24_CAEBL + 88.50 117.70	15.81	466	P17389 human papillomaviru
SwissProt_40:YH24_CAEBL + 88.50 117.22	15.98	490	P31060 escherichia coli.
SwissProt_40:YH24_CAEBL + 88.50 113.31	17.49	739	P15143 gallus gallus (chi
SwissProt_40:YH24_CAEBL + 88.50 108.48	19.55	1228	P55823 bacillus stearoth
SwissProt_40:YH24_CAEBL + 88.50 96.29	25.90	4427	P005470 bacillus subtilis

seq_name: SwissProt_40:NORA_NEIMA

seq_documentation_block:

ID NORA_NEIMA STANDARD; PRT; 447 AA.

AC Q9V0P6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Na(+)-translocating NADH:quinone reductase subunit A (EC 1.6.5.-)
DE (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A) (NQR complex
GN subunit A) (NQR-1 subunit A).
NORA OR NMA0752.

OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 NCBI_taxid=65699;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN-22491 / SEROGROUP A / SEROTYPE 4A;
 RX MEDLINE-20222556; PubMed-10761919;
 RA Parthill J., Achman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S., Jagsis K., Leather S., Moulé S., Mungall K., Quail M.A., Ralstead M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrall B.G.;
 RA "Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491".
 RT Nature 404:502-506(2000).
 RL -1- FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NOR TO NORE ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF UBISEMQUINONE TO UBIQUINOL (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+) (IN) = NAD(+) + UBIQUINOL + NA(+) (OUT).
 CC -1- SUBUNIT: COMPOSED OF SIX SUBUNITS; NOR A, NOR B, NOR C, NOR D AND NOR E (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE NOR A FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC EMBL: AL162754; CAB84035.1;
 CC Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
 KW Complete proteome.
 SQ SEQUENCE 447 AA; 48673 MW; B808CAA82C901AD CRC64;

alignment_scores:
 Quality: 2211.00 Length: 447
 Ratio: 5.059 Gaps: 0
 Percent Similarity: 97.763 Percent Identity: 97.092

alignment_block:
 US-09-303-518D-127 x NOR_A_NEIMA ..

Align seg 1/1 to: NOR_A_NEIMA from: 1 to: 447

```

1 ATGATTAATAATCAAAAAAGCTTAACCTGCCATCGCGGCGACCGGGA 50-
|||||
1 MetIlelysllelysllysllyleuasnleurollealaglyargprogl 17
51 GCAAGTCATTATGACGGCGCGTCATTACGAGAGTGGCGTGGTGGCG 100
|||||
17 uginvalilletyraspelyprovalillethngluvalalaleuenglgs 34
101 AAGAATATGCCGATGCGCGCCCTNGATGAAGTCAAGAGAGCGATGCC 150
|||||
34 luguyluralaglymetarfprosermetlysllyslgylgylaspala 50
151 GTCAAAAAAGCGCAAGTCGCTGTGAGAGCAAAAAGNATCCGGCGCGGT 200
|||||
51 Vallysllyslgylglnvalleuqhgluaspyslyasnproglyvalva 67
201 GTTACCGCGCGCGTTCAGGCAAAATCGCGCATCATCGCGCGGCAAA 250
|||||
67 lphetthralprovalserglylysillealalalehisargllyglul 84
251 AGCGGTAATTCACTGCTGGTGTGATGCGCTTGAAGCAAGCAAGAAATC 300
|||||

```

```

84 ysargvalleuglnservalillealavalgluglyasnaspglulle 100
301 GAGTTCGACCGCTACGGCGCGGAGCGTGGCAACACTTAACGGCGGANGA 350
|||||
101 giuphngluargtyralarogluvalaleuallasnleuserglyglugl 117
351 ATNNNGNNGCAATTCATTCATCCGATCGGTTTGGACTGCGCTGGTANCC 400
|||||
117 uvalargargasnleuileglnserglyleutrtphralaleuargthra 134
401 GTCCGTTACGCAAAATCCCTCGCTGCATGCCGAGCGGTGCCATGTC 450
|||||
134 rpprophserlyslleproalavalaspralagluuprophlealallephe 150
451 GTCAATGGATGAGACCAATCCGCTNCGCGGACACCGCTGGTGTGAT 500
|||||
151 valasnalametaspthrasnproleuallalasprowalvalvalill 167
501 CAAGAAGCGCGNGANGATTCAGACGANGTNGCTGGTATGAGCGGT 550
|||||
167 elysgluallalagluaspheargarglyleuileuvalleuserargl 184
551 TCACCGAGCGTAAATTCATGTGTGTAGGCAAGCTGGCGGACAGTGGCG 600
|||||
184 eutthrgluarglysllehisvalcyslysalalaglyalaspvalapro 200
601 TCTGAATATGCTGCCACATCGAACAACATGATTCGCGCGCGCATCC 650
|||||
201 sergluasalalalalasnilleglutthrhstgubpneglylprohlspr 217
651 GCGCGGTTTGAAGTGGACGCGACATTCATTTCATTGAGCGGTGGTGC 700
|||||
217 oalaglyleuserglythrhstlehisphelleglubprovalaglyalaa 234
701 ACAAACGCTTGGACACATCATTCATCAAGATGTATGATCCATGGAGCT 750
|||||
234 snlysthrvaltrprrhrilleasnlyrglnasprvalillealillelytrg 250
751 TTGTTGCAACAGCGCGCTGTGAACACGAGCGCGGTATGCTTGGGTGG 800
|||||
251 leuphealathrclyargleuasnthrcluargrvalillealaleuglygl 267
801 TTCTCAAGTCAACAACACGCGCTTTCGTAACGCTTGGTGGTGGAAAG 850
|||||
267 ysergluvalasnlysprowargleuileuargthrvallleuglyalatsv 284
851 TATCGCAATTTACTCGCGCGAATGTTGACGCGAGCAACCGCGTGAAT 900
|||||
284 alserglillethralaglyluileuvalaspalaaspasnargvalille 300
901 TCCGCTTGGTATGAACGCGCGGATTCACAAAGCGCGGACGATATTT 950
|||||
301 serglyservalleuasnlyalallethrglnlglalalehisasprtyle 317
951 GGGAGCGTACCAACATCAGATTTCCGTTATGAGAGAGCGCGGCAAG 1000
|||||
317 uglargtyrhlnasnglnlleservalillegluluglyargserlysg 334
1001 ACCTGTGGCGTGGGTGGCGCGGAGCGGACCAATPACTCCATCCAGGCT 1050
|||||
334 luueuqhelytrpvalalaproglmprowaspyslyserillethrarlg 350
1051 ACGACCGTGGCGCATTTCTGAAAAAACAACCTTCAAGATTTCAGACAGC 1100
|||||
351 thrtthleuglyhisheleuylsasnlyslleupheylsphetthrtal 367
1101 CGTCAACGCTGGCGACCGCGCATGTCGCGGATTTGGTACTTACGAGCGG 1150
|||||
367 avalasnlyglylaspargalamevalprolleghlythryrgluargv 384
1151 TATGCGCGTACGACATCCGCTACCGCTTGGCGGATTTAATGCTC 1200
|||||
384 alwetproleuasnilleuuprothleuileuileuargaspheulleval 400

```

seq_name: Sw1ssp07c_40:NR0A_NEIMB

seq_documentation_block:	ID	NORA_NEIMB	STANDARD:	PRT:	447 AA.
AC	Q2KOM3				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Na(+)-translocating NADH-quinone reductase subunit A (NC 1.6.5.-)				
DE	(Na+)-translocating NQR subunit A (Na(+)-NQR subunit A) (NQR complex subunit A) (NQR-1 subunit A).				
GN	NORA OR NM50569.				
OS	Neisseria meningitidis (serogroup B).				
OC	Bacteria: Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.				
OX	NCBI_TaxID=491;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MC58 / SEROGROUP B:				
PX	MEDLINE=20175755; PubMed=10710307;				
RA	Tetteelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J., Nelson W.C., Gwinn M.L., Deboy R., Peterson J.D., Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A., Mason T., Clecko A., Parksey D.S., Blair E., Cifton H., Clark E.B., Cotton M.D., Ullrich T.R., Khouri H., Qin H., Vamathevan J., Gill J., Scariato V., Maignani V., Pizzo M., Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappaport R., Venter J.C.,				
RT	MC58."				
RL	Science 287:1809-1815(2000).				
CC	-1- FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO MORE ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF UBISEMIOQUINONE TO UBIQUINOL (BY SIMILARITY).				
CC	-1- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+) (IN) = NAD(+) + UBIQUINOL + NA(+) (OUT).				
CC	-1- SUBUNIT: COMPOSED OF SIX SUBUNITS; NORA, NORB, NORC, NORD, NORE AND NORE (BY SIMILARITY).				
CC	-1- SIMILARITY: BELONGS TO THE NORA FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement or send an email to license@isb-sib.ch).				
CC	EMBL: AE002412; AAF40997.1; -.				
DR	TIGR: NM00569.				
DR	Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport; Complete proteome.				
KM	SEQUENCE 447 AA; 48635 MW; BE174A78439BE477 CRC64;				
QJ					

```

alignment_block:
US-09-303-518D-127 x NQRA_NEIMB ..
Align seg 1/1 to: NQRA_NEIMB from: 1 to: 447

```

1	ATGATTAAATCAAAAAGGCTTAAACCTGGCCATCGGGGAGACACCGGA	50
1	MetIleuysIleuysIysIsgIyeuAsnIleuProIleIlaIglIyAyrProI	17
51	GCAAGTCATTATAGCAGGGCCGCTCATTTACCAAGTCGCGTTCGTTGGCG	100
17	uGlIAlaValIyAspOIProIAlaIlethrGluValAlaIleuIsgIyG	34
101	AAGAAATGCGCGGATGACCGCCCTTNCATGAAGTCAGGAAGCGCATGCC	150
34	IuGIuIyTAlaIglIyMeIyAyrProSeMeIlyIyValIySdIuOIyAspIa	50
151	GTCAAAAAGGCCAAGTCTGTTTGAAGACAAAAGANATCCGGCGGTGGT	200
51	VallIyIysIsgIyInValIleuPheGluAspIyIysAsnProGIyValIyA	67
201	GTTTACCGCCCGGTTTACGAGCAAAATCCGGCCATTCATCGCGGCGAA	250
67	IpheThrIalProIAlaSerOIyIysIleIlaIAlaIleIhaIsgIyIuI	84
251	AGCGCGTACTTCAGTGGGTGTGATTCGCGTGAAGGACAGCAGCAATC	300
84	YsaIyAlValIeuGInSerValIleIlaIValIglIyAsnAspIuIle	100
301	GAGTTCACAGCGTACCGCGCCGGAAGGTTGGCAAACTTAACGGCGANGA	350
101	GluPheIuIygrIyThrIalProGIuAlaIeuAlaAsnIleuSerGIyGluI	117
351	ANTNKGNGCAATCTATCCATCCGGTTCGTGACATCGCTGCGGTANCC	400
117	uValIyAlyArAsnIleuIleOIserOIyeuThrPhrIAlaIeuArGThA	134
401	GTCGGTTACACAAAATCCCTCGCGGTGATGCGGACCGCTGCGCATCTC	450
134	rgrProPheSerIyIleProIAlaValAspAlaIgluIProPheAlaIlePhe	150
451	GTCAATGCGATGGACACCAATCCGCTNGCGGACAGCCGTGTGGTGTAT	500
151	ValAsnIaMeIasPhrIasnProIeuAlaIaIAspProIValIleI	167
501	CAAGAAGCGCGNGANGATTTTACAGCGANGTGGTGTATTGACCGCT	550
167	eIyIsgIuAlaIaIglIuAspPheIyArGIyIleuIeuValIleuSerArGIyI	184
551	TGACCGGAGGTAAATTCATGTGTGTAAAGCGACGTGGCGGACGTCGCG	600
184	euThrGIuArGIyIleIhaIysValIySylAlaIaIaIglIyAlaAspValPro	200
601	TTCGAAAATCTCTCCAAATTCGAAACACATGAATTCGGCGCGCGGCATCC	650
201	SeIcIuAsnAlaIaAsnIleOIuThrIasGIuPheGIyIyProIAspI	217
651	GCGCGGTTTGAGTGGCAGCAGCATTCATTTCATTGAGCGCGTCTGTCGAA	700
217	oAlaIglIyIeuSerGIyThrIhaIleIhaIAspHeIeOIuProValIglIyAla	734
701	ACAAAACCGGTTGAGCATTCATTTACAAAGTAAATTCGCATCGGAGCGT	750
234	snIyIshThrValItrPhrIleAsnIyTGIuAspValIleThrIleGIyArG	250
751	TTCGTTTGCAACAGGCGGTGTGAACACGAGCGCGTGTATGCTGTGGGTGG	800
251	IeuPheAlaIthrGIyArGIyIeuAsnThrGIuArGIyAlIleAlaIeuIsgIyI	267
801	TTCGCAAGTCACAAAACAGCGCTTCGCAACCGTTTGGGTGCGAAG	850
267	ySerGIuIuValAsnIySPProGIuIeuIaIuGIyThrValIeuOIyAlaIyIyS	284

```

851 TATGCAATTACTGGGGGAATGGTTGACGACGACGCGGATT 900
|||||
284 aISerGlnIleThrAlaGlyLeuValAspThrAspAsnArgValIle 300
|||||
901 TCCGCTGGCTATTGAAGCGCGGATTACACAGCGCGCGCATTAATT 950
|||||
301 SerGlySerValLeuAsnGlyAlaIleThrGlnGlyAlaHisAspTyrIle 317
|||||
951 GGGACGCTACCAATGATGATTCCTGTTATGCAAGAGCGCGCAAG 1000
|||||
317 uGlyArgTyrHisAsnGlnIleSerValIleGlnGlnGlyArgSerLysG 334
|||||
1001 AGCGTTGGCTGGGTGGCGCGCGCGGCAAAATGCTCCACGCGCT 1050
|||||
334 IuLeuPheGlyTyrPValAlaIleProGlnProAspLysTyrSerIleThrArg 350
|||||
1051 ACGACCTCGGCGCATTTCTGAAAAACAACTCTCAAGTTACAGCAGC 1100
|||||
351 ThrThrLeuGlnHisPheLeuLysAsnLysLeuPheLysPheAsnThrAl 367
|||||
1101 CGTAAAGGTGGGACCGCGCGCATGGTGGCGGATTGTTACTTACAGCGCG 1150
|||||
367 aValAsnGlyGlyAspArgAlaMetValProIleGlyThrTyrGluArgV 384
|||||
1151 TAAATCGCGTAAAGACATCTGCTACCTGCTTTGGCGGATTAAAGCTC 1200
|||||
384 aIMetCProLeuAspIleLeuProThrIleuLeuAsnArgAspIleVal 400
|||||
1201 GGCATACCGACGACGCGCAAGCATTTGGTGTGGAATTGACGAGA 1250
|||||
401 GlyAspThrAspSerAlaGlnAlaLeuGlyCysLeuGlnIleuAspGlnG 417
|||||
1251 AGACCTGGCTTTGCGACGCTTCGCTCGCCGCGGCAATAGCAATANGCC 1300
|||||
417 uAspLeuAlaLeuLysSerPheValCysProGlyLysTyrGluTyrGlyP 434
|||||
1301 CGCTGTGGCTAAAGTGTGGAACNTTGAAGAAAGAGAGC 1341
|||||
434 rOleuLeuArgLysValLeuGlnThrIleGlnLysGlnGly 447
|||||
seq_name: SwissProt_40:NORA_PASMU
seq_documentation_block:
ID NORA_PASMU STANDARD; PRT; 446 AA.
AC
O9CUB1:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE (Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE (Na(+)-translocating NOR subunit A) (Na(+)-NOR subunit A) (NOR complex
DE subunit A) (NOR-1 subunit A).
GN NORA OR PM1328
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whitam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida PM70."
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC - FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO MORE
CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC UBISEMIQUINONE TO UBIQUINOL (BY SIMILARITY).
CC - CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) +
CC UBIQUINOL + NA(+)(OUT).
CC - SUBUNIT: COMPOSED OF SIX SUBUNITS; NORA, NORB, NORC, NORD, NORE
CC AND NORF (BY SIMILARITY).

```

```

CC -1- SIMILARITY: BELONGS TO THE NORA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE006171; AK03412.1;
KW Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
KW Complete proteome.
SQ SEQUENCE 446 AA; 48168 MW; 5C5287CB77C4932D CRC64;

alignment_scores:
Quality: 1614.50 Length: 447
Ratio: 4.087 Gaps: 1
Percent Similarity: 88.367 Percent Identity: 69.351

alignment_block:
US-09-303-518d-127 x NORA_PASMU
Align seg 1/1 to: NORA_PASMU from: 1 to: 446

1 ATGATTAATAATCAAAAAAGCTTAACCTGCCCATCGCGGACAGACGGA 50
|||||
1 MetIleThrIleLysGlyGlyLeuAsnLeuProIleSerGlySerProG 17
51 GCAAGTCATATTATGACGGGCGCGCATTAACCGAAGTGGCTGCTGGCG 100
|||||
17 uGlnValIleArgAspGlyAsnAlaIleThrGlnValAlaLeuLeuGlyG 34
101 AAGATATGCGCGGTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 150
|||||
34 IuGlnTyrValAlaGlyMetArgProSerMetLysValArgGlnGlyAspVal 50
151 GTCAAAAAAGGCCAAGTGTGTTGAAGACAAAGAAAGNATCCGGCGGTG 200
|||||
51 ValLysLysGlyGlnValLeuPheGlnAspLysLysAsnProGlyValVa 67
201 GTTTACCGCGCGCGGTTCAGGCAAAATCGCGCATTCATGCGGCGCAA 250
|||||
67 IPhetThrAlaProAlaSerGlyThrValThrAlaIleHisArgGlyAlaL 84
251 AGCGGTACTTCAGTGCAGTGCATGATGCGGCGGCGGCGGCGGCGGATC 300
|||||
84 yAsArgValLeuGlnSerValProAlaValAspAlaThrProValSerIlePhe 150
301 GAGTTGCAAGCGTACGCGCGCGGCGGCGGCGGCGGCGGCGGCGGANGA 350
|||||
101 ThrPheGlnLysTyrThrThrGlnIleuAsnGlnIleuThrSerGlnG 117
351 ANTNNGNNGCATCTGATCCATCCGTTTGGGACTGGCGGCGGTANCC 400
|||||
117 nValArgGlnAsnLeuGlnAlaSerGlyLeuTyrPThrAlaLeuArgTrrA 134
401 GTCCGTTTCAGAAAAATCCCTGCGGTCGATGCGGCGCGGCGGCGCATCTTC 450
|||||
134 rGProPheSerLysValProAlaValAspAlaThrProValSerIlePhe 150
451 GTCAATGCGATGACCAACCAATCCGCTGCGGCGGCGGCGGCGGCGGAT 500
|||||
151 ValAsnAlaMetAspThrAsnProLeuLysAlaAspProGlnValIleVa 167
501 CAAAGAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 550
|||||
167 IeGlnIleSerAlaGlnAlaPheGlnAlaGlyLeuThrValLeuSerArgV 184
551 TGACCGAGCGGTAATCATGTGTGTAAGCAGCGGCGGCGGCGGCGGCGG 600
|||||
184 euHisGlnGlyLysValTyrLeuGlyLysAlaAlaAsnLaserIlePro 200

```

OS Haemophilus influenzae.
OC Bacteria: Proteobacteria; gamma subdivision: Pasteurellaceae.
OC Haemophilus.
OX NCBI_TaxID=727;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,
Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
Utterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
Fline L.D., Fritchman J.L., Fuhrtman J.L., Geoghagen N.S.M.,
Rau Grehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
[2]
RP REVISIONS.
RA White O., Clayton R.A., Kerlavage A.R., Fleischmann R.D.;
RL Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.
[3]
RP SEQUENCE OF 1-75 FROM N.A.
RC STRAIN-RD / KW20 / ATCC 51907;
RA Barcak G.J., Helmer S.R.;
RL Submitted (JAN-1995) to the EMBL/Genbank/DBJ databases.
[4]
RP IDENTIFICATION AS NOR SYSTEM.
RC STRAIN-RD / KW20 / ATCC 51907;
RX MEDLINE=96176316; PubMed=8601449;
RA Hayashi M., Nakayama Y., Umemoto T.;
RT "Existence of Na⁺-translocating NADH:quinone reductase in Haemophilus
influenzae.";
RL FEBS Lett. 361:174-176(1996).
[5]
RN IDENTIFICATION BY MASS SPECTROMETRY.
RP MEDLINE=20137486; PubMed=10675023;
RX Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wlpi B.,
RA Gray C., Fountoulakis M.;
RL "Two-dimensional map of the proteome of Haemophilus influenzae.";
RT Electrophoresis 21:411-429(2000).
[6]
CC -I- FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NOR TO NOR
ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
UBIQUINONE TO UBIQUINOL (BY SIMILARITY).
CC -I- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) +
UBIQUINOL + NA(+)(OUT).
CC -I- SUBUNIT: COMPOSED OF SIX SUBUNITS: NOR A, NOR B, NOR C, NOR D,
AND NOR E (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE NOR A FAMILY.
CC -I- CAUTION: REP.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
FRAMESHIFT IN POSITION 63.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32702; AAC21836.1; -;
DR EMBL: U20229; AAA62134.1; ALT_FRAME.
DR TIGR: H10164; -;
KW Oxidoreductase; NAD; ubiquinone; Transport; Sodium transport;
KW Complete proteome.
SQ SEQUENCE 447 AA; 48884 MW; 4670C93FB5FF0912 CRC64;

alignment_scores:

Quality: 1574.00 Length: 448
 Ratio: 4.057 Gaps: 2
 Percent Similarity: 86.607 Percent Identity: 67.634

alignment_block:

US-09-303-518D-127 x NORA_HAEIN

Align seg 1/1 to: NORA_HAEIN from: 1 to: 447

```

1 ATGATTAATAAATAAAGCTTAACCTGCCATCGCGGACAGCCGGA 50
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 MetIleThrIleLysGlyLeuAspLeuProIleAlaGlyLysProAl 17
51 GCAAGTCATTATGACGGGCCCATTAACGAGTGGCGTTGGTGGCG 100
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 agInValIleHisSerGlyAsnAlaValAsnGlnValAlaIleLeuGly 34
101 AAGAATATGCGGATATGCGCCCTNGATGAAGTCAGAGGAGCGGATGCC 150
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 IuGlutylValGlyMetArgProSerMetLysValArgGluGlyAspVal 50
151 GTCAAAAAAGCCCAAGTCTGTGTAAGACAAAAAGNATCCGGCGGTGCT 200
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 ValLysLysGlyGlnValLeuPheGluAspLysAsnProGlyValAl 67
201 GTTTCACCGCGCGNGTTTCAGGCAAAATGCGCCCATCGCGGCGGAA 250
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
67 ePheThrAlaProAlaSerGlyThrIleThrAlaIleAsnArgLysGlu 84
251 AGCGGCTACTTCAGTCGGTGTGATGCGCTGTAAGGCAAGCAAGCAATTC 300
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
84 ysArgValLeuGlnSerValValIleAsnValGluGlyAspGluLysIle 100
301 GAGTTCGAACGCTACGCCCGGAGCGCTGGCAAACTTAAGCGCGGANGA 350
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 ThrPheAlaLysTyrSerThrGlnIleLeuAsnThrLeuSerSerGlu 117
351 ANTNNNGNCAATCTGATCCAATCCGTTTGTGACTGCTGCTGCTANCC 400
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 nValLysGlnAsnLeuIleGlnSerGlyLeuThrIleAlaLeuArgThr 134
401 GTCCGTTACAGCAAAATCCCTGCGTGCATGCCAGCGCTTGCATCTTC 450
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134 rgProPheSerLysValProSerIleGlnSerGluAlaSerSerIlePhe 150
451 GTCAATGGGATGACCAATCCGCTNGCGAGACCTCGTGGTGTGAT 500
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProSerValVal 167
501 CAAGAAGCCGCGNGANGATTTCAGACGANGTNGCTGATTTGAGCCGTT 550
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167 uLysGluTyrSerGlnAspPheThrAsnGlyLeuThrValLeuSerArg 184
551 TGACCGAGCGGTAA..ATCCATGTGTGTAAAGCAGCTGGCGGACAGCTG 597
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
184 euPheProSerLysProLeuHisLeuGlyLysAlaGlyAspSerAsnIle 200
598 CGGTGGAATAATGTCGCAACATGCAACATCAATGATCGGCGGCCGCA 647
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
201 ProThrAlaAspLeuGlnAsnLeuGlnIleHisAspPheThrGlyVal 217
648 TCCGGCCGCTTGTAGTGCACGACATTCATTCATTCATGAGCCGCTG 697
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
217 sProAlaGlyLeuValGlyThrHisIleHisPheIleAspProValAl 234
698 CAACAAAAACGTTTGGACCATCAATATATCAAGATGTAATGCCATGGA 747
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
234 IeGlnIleThrValIleThrHisIleAsnTyrGlnAspValAlaIleVal 250
748 CGTTTGTGTTGCAACAGCGCGTGTGAACACCGCGCGATGCTTGGG 797
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
251 LysLeuPheThrThrGlyGlnLeuTyrSerGlnArgValIleSerLeuAl 267

```

```

798 TGCTCTCAAGTCACAAACACGACGCTTGTGCTACCGCTTTGGGTGCGA 847
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
267 agLysProGlnValLysGluProArgLeuValArgThrIleGlyAlaAl 284
848 AAGTATGCAAAATTACTGCGGCGGAATGTTGACGACAGCAACCGCGTG 897
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
284 snLeuSerGlnLeuThrGlnAsnGlnLeuSerAlaGlyLysAsnArgVal 300
898 ATTTCCGGTTCGTTATGAACGGCGGATTACACAAAGCGCGCAGATTA 947
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301 IleSerGlySerValLeuGlyGlnIleAlaLysAspSerHisAspTyr 317
948 TTTGGACGCTACACCAATCAAGATTTCGCTATTCGAAGGCGCCAGCA 997
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
317 rLeuGlyArgTyrAlaLeuGlnValSerValIleAlaGlnGlyAsnGlu 334
998 AAGAGCTGTTGCGCTGGTTGGCGCCGACCGGACAAATATTCATCAGC 1047
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
334 ysgIlePhePheGlyTyrIleMetProGlnAlaAsnLysTyrSerValThr 350
1048 CGTAGCACCGCTGCGCCATTTCTGAAAAACAAACTCTCAAGTTCAGCAGC 1097
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
351 ArgThrValLeuGlnLysHisPheSerLys...LysLeuPheAsnPheThr 366
1098 AGCCGTCAACGGTGGCGACCGCGCCATGTCGCGATTTGCTACTACGAGC 1147
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
366 rSerGluAsnGlyGlyGluArgAlaMetValProIleGlySerTyrGlu 383
1148 GCGTAATCCGCGTACAGCATCTGCTACCTGCTGTTGGCGCATTTAATC 1197
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
383 rGlyAlaMetProLeuAspIleLeuProThrLeuLeuLeuArgAspLeuIle 399
1198 GTGCGGATATACGACAGCGCGCAAGCATTTGGTGTGCTGTAATGAGCA 1247
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
400 ValGlyAspThrAspGlyAlaGlnGlnLeuGlyCysLeuGlnLeuAsp 416
1248 AGAAGACCTGCTGTTTGTGACGCTTGTGCGCCGCGCAATACGAATTANG 1297
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
416 uGlnAspLeuAlaLeuGlySerPheValCysProGlyLysTyrGluTyrG 433
1298 GCCCGCTGTTGCGTAGGTGCAAAACNTGAGAGGAAGGAC 1341
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
433 LysIleLeuArgGlnValLeuAspLysIleGlnLysGluGly 447
seq_name: SwissProt_40:NORA_VIBHA
seq_documentation_block:
ID NORA_VIBHA STANDARD PRT: 446 AA.
AC Q9RFW1;
DT 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE (Na(+))-translocating NQR subunit A (Na(+))-NQR subunit A (NQR complex
DE subunit A) (NQR-1 subunit A).
OS NORA.
GN Vibrio harveyi.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=669;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BB120;
RX MEDLINE=20056044; PubMed=10587447;
RA Zhou W., Bertsova Y.V., Feng B., Tsatsos P., Verkhovskaya M.L.,
RA Genis R.B., Bogachev A.V., Barginera B.;
RT "Sequencing and preliminary characterization of the Na+-translocating
RT NADH:ubiquinone oxidoreductase from Vibrio harveyi.";
RL Biochemistry 38:16246-16252(1999).
CC -!- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO NQRE
CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF

```


DE Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
 DE (Na(+)-translocating NADH-quinone reductase subunit alpha) (Na(+)-
 DE translocating NOR subunit A) (Na(+)-NOR subunit A) (NOR complex
 DE subunit A) (NOR-1 subunit A).
 GN NORA OR NOR1.
 OS *Vibrio alginolyticus*.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; *Vibrio*.
 OX NCBI_TaxID=663;
 RN [1]
 RN SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
 RP STRAIN-NCIMB 11038;
 RX MEDLINE=95104445; PubMed=7805867;
 RT Beattie P., Tan K., Bourne R.M., Leach D.R.F., Rich P.R., Ward F.B.;
 RT "Cloning and sequencing of four structural genes for the Na(+)-
 RT translocating NADH-ubiquinone oxidoreductase of *Vibrio*
 RT *alginolyticus*.";
 RT FEBS Lett. 356:333-338(1994).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP Hayashi M., Unemoto T., Sugiyama A.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN SEQUENCE OF 1-10.
 RP MEDLINE=98149659; PubMed=9490015;
 RA Nakayama Y., Hayashi M., Unemoto T.;
 RT "Identification of six subunits constituting Na⁺-translocating NADH-
 RT quinone reductase from the marine *Vibrio alginolyticus*.";
 RT FEBS Lett. 422:240-242(1998).
 RN [4]
 RN SEQUENCE OF 1-9 AND 334-340.
 RP MEDLINE=95104444; PubMed=7805866;
 RA Hayashi M., Hirai K., Unemoto T.;
 RT "Cloning of the Na(+)-translocating NADH-quinone reductase gene from
 RT the marine bacterium *Vibrio alginolyticus* and the expression of the
 RT beta-subunit in *Escherichia coli*.";
 RT FEBS Lett. 356:330-332(1994).
 RN [5]
 RN INHIBITION OF ENZYMATIC ACTIVITY.
 RP MEDLINE=20016049; PubMed=10549856;
 RA Nakayama Y., Hayashi M., Yoshikawa K., Mochida K., Unemoto T.;
 RT "Inhibitor studies of a new antibiotic, koromicin, 2-n-heptyl-4-
 RT hydroxyquinoline N-oxide and Ag⁺ toward the Na⁺-translocating NADH-
 RT quinone reductase from the marine *Vibrio alginolyticus*.";
 RT Biol. Pharm. Bull. 22:1064-1067(1999).
 RN [6]
 RN REVIEW.
 RP MEDLINE=21145117; PubMed=11248187;
 RA Hayashi M., Nakayama Y., Unemoto T.;
 RT "Recent progress in the Na(+)-translocating NADH-quinone reductase
 RT from the marine *Vibrio alginolyticus*.";
 RL Biochim. Biophys. Acta 1505:37-44(2001).
 RN [7]
 RN REVIEW.
 RP MEDLINE=21145118; PubMed=11248188;
 RA Steuber J.;
 RT "Na(+)-translocation by bacterial NADH:quinone oxidoreductases: an
 RT extension to the complex-I family of primary redox pumps.";
 RL Biochim. Biophys. Acta 1505:45-56(2001).
 CC -1- FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
 CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
 CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO NORE
 CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
 CC UBISEMIOQUINONE TO UBIQUINOL.
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+) (IN) = NAD(+) +
 CC UBIQUINOL + NA(+) (OUT).
 CC -1- ENZYME REGULATION: THIS REACTION IS TIGHTLY COUPLED TO THE NA(+) +
 CC PUMPING ACTIVITY AND SPECIFICALLY REQUIRES NA(+) FOR ACTIVITY.
 CC INHIBITED BY KOROMICIN AND 2-N-HEPTYL-4-HYDROXYQUINOLINE N-OXIDE
 CC (HONO).
 CC -1- SUBUNIT: COMPOSED OF SIX SUBUNITS, NORA, NORB, NORC, NORD, NORE
 CC AND NORE.
 CC -1- SIMILARITY: BELONGS TO THE NORA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z37111; CAA85476.1; -
 KW EMBL: AB008030; BAA22910.1; -
 DR Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport.
 FT CONFLICT 337 337 W -> L (IN REF. 4).
 SQ SEQUENCE 446 AA; 48622 MW; 6D65ACAA53FE515C CRO64;
 alignment_scores:
 Quality: 1422.50 Length: 447
 Ratio: 3.763 Gaps: 1
 Percent Similarity: 84.564 Percent Identity: 61.074
 alignment_block:
 US-09-303-518D-127 x NORA_VIBAL ..
 Align seg 1/1 to: NORA_VIBAL from: 1 to: 446
 1 ATGATTAAATCAAAAAGGCTTAACCTGCCCATCGCGCGGACGACCGA 50
 1 Metilethielyslyslsleuasprouleialaglythnprose 17
 51 GCAAGTCATTTATGACGGCCGCTCATTTACCGAAGTCGGTGTGGCG 100
 17 rglnvalilleasnaspilylsthlelyslsyalalaleuaglyg 34
 101 AAGAATATGGCGGTATGCGCCCTTGATGAAGTCAAGGAGGCGATGCC 150
 34 lndluyrvalglgmetargprothmethisvalargvalglaspilu 50
 151 GTCAAAAAGCCCAAGTGTGTTGAAGACAAAAGNATCGGGCGTGT 200
 51 Vallslsyalaglnvalleuphegluasplslsyalasproglyally 67
 201 GTTTACCGCGCCGTTTCAGGCAAAATCCGCCGATCCATCCGCGGAGA 250
 67 spherthlalaalaglylslsvalillegluvalasnaraglyalal 84
 251 AGCGGCTACTCATGCGTGTGATTCGCTTGAAGCAAGCAAGCAATC 300
 84 ysargyvalleuglnservalvalillegluvalalaglylgluval 100
 301 GAGTTGCAACGCTACGCGCGCGAAGCGTTGCAAACTTAAGCGCGANGA 350
 101 Thrpheaspyspsheluaalaglnleuserglyleuaspargluval 117
 351 ANTNNGNNGCAATCTGATCCATCCGTTGTGATCGCGCGCTGATC 400
 117 lilelythglnleuvalaspseryleuylprthlaleuarglytha 134
 401 GTCCGTCGACCAAAATCCGCGCGATCGATCGAGCGGTTCGCAATTC 450
 134 rgtpropheserlyvalproalillegluserthrlysalalephe 150
 451 GTCAATCGATGACCAACCAATCCGCTNGCGGAGACCGCTGTGTGAT 500
 151 Valthrhlamelasphrhnasproleualalalyprogluuleuilel 167
 501 CAAGAAGCCGCGANGATTTGACGACGANGTNGCTGGTATTAGCGGTT 550
 167 easnglglnglgluvalahelealaglyleuasplileleuseralal 184
 551 TGACCGAGCGTAAATCATGTGTGTAAGCAGCAGTGGCGGAGCGCG 600
 184 euhthglnglylslsvaltyvalcylsls...serglythrserleupro 199


```

601 TGTGAAATGCTCCACATCGAAGACATGATTCGGCGCCGACATCC 650
      ::::: ::::: ||| ||| ||| ||| ||| ||| |||
200 ArgSerSerGlnSerAsnValGlnGluHisValPheAspGlyProHispr 216
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
651 GGGCGGTTGAGTGCGACGACATTCATTCATTCAGCCGCGTGGTGA 700
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
216 GAAAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 233
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
701 ACAAACGCTTGGACATCATTCATTCATTCATTCATTCATTCATTC 750
      ::::: ::::: ||| ||| ||| ||| ||| ||| ||| |||
233 LAsnValAlaTrpSerIleAsnTrpGlnAspValIleAlaPheGlyLys 249
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
751 TTGTTGCAACAGCGCGTGTGAACAGCGCGCGTGTGCTTGGTGG 800
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
250 LeuPheLeuThrGlyLeuLeuTrpAspArgValValSerLeuAlaG 266
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
801 TTCTCAAGTCAACAACCGCGCTTGGTGTGCGTGTGCGTGTGCG 850
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
266 YProValValAsnProAlaGlyLeuValArgThrValIleGlyAlaSer 283
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
851 TATCGCAATTTACTGCGGCGAATTTGTTGACGACAGACACCGCG 900
      ::::: ::::: ||| ||| ||| ||| ||| ||| ||| |||
283 euAspAspLeuThrAspAsnGluLeuMetProGlyGluValArgValIle 299
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
901 TCCGTTGCGTATTGAACGCGCGATTAACAGCGCGCGCGATTTAT 950
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
300 SerGlySerValLeuThrGlyThrHisAlaThrGlyProHisAlaTrp 316
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
951 GGGACGCGTACCAATCAGATTCGTTGTTGAGAGAGCGCGCGCAAG 1000
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
316 uGlyArgGlyHisGlnGlnValSerValLeuArgGlyArgGlyLysG 333
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1001 AGCTGTTGCGGTTGCGGCGCGCGCGCGCGCAATTCATTCACGCG 1050
      ::::: ::::: ||| ||| ||| ||| ||| ||| ||| |||
333 LulePheGlyTrpAlaMetProGlyLysAsnLysPheSerValThrArg 349
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1051 ACAGACCTGCGCGCATTCCTGGAAGAAACAACTCTCAAGTTACAG 1100
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
350 SerPheLeuGlyHisValPheLysGlyGlnLeuPheAsnMetThrTrh 366
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1101 CGTCAACGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1150
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
366 rThAsnGlySerAspArgSerMetValProIleGlyAsnTrpGlyLys 383
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1151 TAATGCGCGTACAGATTCCTGCTACCTGCTTGGCGCGATTAATG 1200
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
383 AlMetProLeuAspMetGluProThrLeuLeuArgAspLeuGlyAla 399
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1201 GGGCGTACCGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1250
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
400 GlysAspThrAspSerAlaGlnAlaLeuGlyAlaLeuGlnLeuAspG 416
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1251 AGACCTGCTTGTGAGTCTGCTGCTGCGCGCGCGCGCGCGCGCG 1300
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
416 uAspLeuAlaLeuGlyThrPheValLysProGlyLysTrpGlyLysG 433
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1301 CGCTGTTGCGTAAAGTCTGGAACCTTGAAGAAGAGCG 1341
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
433 hrLeuLeuArgGlyLysLeuAspThrIleGlyLysGlyLys 446
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
seq_name: swissprot_40:NORA_VIBCH

```

```

seq_documentation_block:
ID NORA_VIBCH STANDARD: PRT: 446 AA.
AC 09KPS1; 09X4Q3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE (Na(+)-translocating NADH-quinone reductase subunit A) (NOR complex
DE subunit A) (NOR-1 subunit A).
GN NORA OR VC2295.
OS Vibrio cholerae.

```

```

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxId=666.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=99179036; PubMed=10077658;
RA Haese C.C., Mekalanos J.J.
RT "Effects of changes in membrane sodium flux on virulence gene
RT expression in Vibrio cholerae."
RL Proc. Natl. Acad. Sci. U.S.A. 96:3183-3187(1999).
RM [2]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.T.,
RA Dodsan R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragol I., Sellers P.,
RA McDonauld L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
CC -!- FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO NORE
CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC URISEMIQUINONE TO UBIQUINOL.
CC -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+) (IN) = NAD(+) +
CC UBIQUINOL + NA(+) (OUT).
CC -!- SUBUNIT: COMPOSED OF SIX SUBUNITS; NORA, NORB, NORD, NORE
CC AND NOLF (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NORA FAMILY.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AF117331; AAD29962.1;
DR EMBL: AE004300; AAP95439.1; ALT_INT.
DR TIGR: VC2295;
KW Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
KW Complete proteome.
SO SEQUENCE 446 AA; 48624 MW; 428E8C397EBA163D CRC64;

```

```

alignment_scores:
Quality: 1412.50 Length: 447
Ratio: 3.767 Gaps: 1
Percent similarity: 83.893 Percent identity: 60.403

```

alignment_block:

us-09-303-518d-127 x NORA_VIBCH

Align seg 1/1 to: NORA_VIBCH from: 1 to: 446

```

1 ATGATTAATCAAAAAAGGTCTAAACCTGCCCGGCGGCGAGACGGA 50
||||| ||||||||| ||||||||| ||||||||| |||
1 MetIleThrIleLysGlyLysLeuAspLeuProIleAlaGlyThrProse 17
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
51 GCAAGTCATTTATGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 100
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
17 GlnValIleSerAspGlyLysAlaIleLysValAlaLeuLeuGlyG 34
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
101 AAGATATGCGCGCGTATGCGCGCGCGCGCGCGCGCGCGCGCG 150
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
34 LngLysTrpValGlyMetArgProThrMetHisValArgValGlyAspGln 50

```

```

151 GTCAAAAAGCCAGTGTCTTTGAAGACAAAAGNATCCGGCGTGGT 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 VallysYsAlaGlnIleuPheGluAspLysAsnProGlyVally 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 GTTACCGCGCCGTTTACGACAAATCGCCGATCCATCGCGCGAAA 250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 sPheThSerProValSerGlyValValGluIleAsnArgGlyAla 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
251 AGCCGCTACTCAGTGGTGTATTCCTGAGGACGACGACGAAATC 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 ySaTValIleuGlnSerValValIleGluValAlaGlyAspAspGlnVal 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301 GAGTTCAGACGCTACGCGCGGAGCGCTTGCAAACTTAAGCGCGGANGA 350
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 ThrPheSpsYrPheGluAlaAsnGlnIleuAlaSerLeuAsnArgAspAl 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
351 ANINNGNCANATGTATCCATCCGCTTTGTGACTGCGCTGCGRANC 400
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 alIeLysThrGlnIleuValGlnSerGlyLeuTyrThrAlaPheArgThra 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
401 GTCCGTCAGCAAAATCCCTGCTGATGCGGAGCGCTTGCGCATTC 450
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
134 rGProPheSerLysValProAlaIleAspSerThrSerGluAlaIlePhe 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
451 GTCATGCGATGAGACACCAATCCGCTNGCGGACGACCCCTGTGTGTAT 500
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 ValThrIleMetAspThrAsnProLeuAlaGluProThrValValIle 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
501 CAAAGAACCGCGANGATTTACAGACGANGTNGCTGTATTGACCGGTT 550
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 eaSnGlnGlnSerGluAlaPheValAlaGlyLeuAspValLeuSerAla 184
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
551 TGACGAGCGCTAAATCCATGCTGTGTAGGAGAGCGGACGAGAGTCCG 600
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
184 eunThrGlyLysValTyrValGlyLys..LysGlyThrSerLeuPro 199
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
601 TCTGAAATGCTGCCAATCGAACAATGAAATTCGCGCGCGCATCC 650
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
200 ArgSerGlnGlnProAsnValGlnGlnIleValPheAspGlyProHisPr 216
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
651 GCGCGGTTGAGTGGACGACGACATTCATTATGACCGCGTGGTGCAG 700
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
216 oAlaGlyLeuAlaGlyThrHisMetHisPheLeuTyrProValSerAla 233
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
701 ACAAAACGCTGGGACCATATATCAAGATGAATTCATCCATGGAGCGT 750
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
233 sPhIAsValAlaTyrSerIleAsnTyrGlnAspValIleAlaValGlyGln 249
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
751 TTGTTTGCACAGCGCGCTGTGACACGCGCGGTGATTGGGTGG- 800
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
250 LeuPheLeuThrGlyLeuLeuTyrThrGlnArgValValSerLeuAlaGly 266
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
801 TTCTCAAGTCAACAACACCGCCTTTCGCGTGGTGGTGGTGCAGAG 850
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
266 yProValValAsnLysProArgLeuValArgThrValMetIleValSerL 283
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
851 TATCGCAATTAATCTGCGGCGAATGTTGAGCGACACACCGCGTAT 900
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
283 euGlnGlnIleuValAspSerGluIleMetProGlyGluValArgIleLe 299
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
901 TCCGGTTCGATTGACGCGCGATTAACACAGCGCGCGCATGATTATT 950
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
300 SerGlySerValLeuSerGlyThrLysAlaThrGlyProHisAlaTyrLe 316
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
951 GGGAGCGTACACATCAGATTCCGTTATGAGAGAGCGCGCGCAAG 1000
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
316 uGlyArgThrHisLeuGlnValSerValLeuArgGlnGlyArgAspLysG 333
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1001 AGCTTTGGCTGGGTTGCGCGACGCGCAATAATCTCCATCAGCGGT 1050
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
333 IuLeuPheGlyTyrAlaMetProGlyLysAsnLysPheSerValThrArg 349
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1051 ACGACCGTCCGCGCATTTCTGAAAAACAACCTTTCAAGTTCCAGACAGC 1100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

350 SerPheLeuGlyHisLeuPheLysGlyGlnValTyrAsnMetThrThr 366
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1101 CGTCAAGGTGGCGCGCGCGCATGTCGATTCATTCAGACGCGC 1150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
366 rThrsnGlySerAspArgSerMetValProIleGlyAsnTyrGlnLysV 383
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1151 TAATGCGCTAGACATCCCTACCCGCTTTTGCGGATTTAATGCTC 1200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
383 alMetProLeuAspMetGluProThrLeuLeuArgAspLeuLysAla 399
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1201 GCGCATACCGACGCGCGCGCATTTGGTTCCTTGGAAATGAGCAAGA 1250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
400 GlyAspSerAspSerAlaValArgLeuGlnLysLeuLeuAspGluG 416
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1251 AGACCTCGCTTGTGACGCTGCTGCGCGGCGCAATATGAAATGAGCC 1300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
416 uSpLeuAlaLeuCysTyrPheValCysProGlyLysTyrGlnLysG 433
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1301 CGCTGTTCGTAGAGTGTGGAACCTTGAAGAGGAGC 1341
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
433 IuLeuLeuArgGluCysLeuAspLysIleGluLysGluGly 446
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: SwissProt.40:NORA_PSEAE

seq_documentation_block:

```

ID NORA_PSEAE STANDARD; PRT; 445 AA.
AC O9H2K6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE (Na(+)-translocating NOR subunit A) (Na(+)-NOR subunit A) (NOR complex
DE subunit A) (NOR-1 subunit A).
DE NORA OR PA2999.
GN Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
CC -!- FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO MORE
CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC UBISEMIQUINONE TO UBIQUINOL (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+) (IN) = NAD(+) +
CC UBIQUINOL + NA(+) (OUT).
CC -!- SUBUNIT: COMPOSED OF SIX SUBUNITS: NORA, NORB, NORC, NORD, NORE
CC AND NORF (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NORA FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AF004724; AAC06387.1; -.
CC Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
KW

```

KW Complete proteome.
SQ SEQUENCE 445 AA; 48081 MW; ED0468D06559B701 CRC64;

alignment_scores:

Quality: 1298.00 Length: 447
Ratio: 3.576 Gaps: 1
Percent Similarity: 81.208 Percent Identity: 57.271

alignment_block:

US-09-303-518D-127 x NORA_PSEAE

Align seg 1/1 to: NORA_PSEAE from: 1 to: 445

```

1 ATGATTAATAATCAAAAAGGCTTAACCTGGCCCATCGGGGCGAGACGCGGA 50
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 MetIleuysIleuysArgGlyLeuaspLeuProIleSerGlyAlaProG1 17
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 uGlnArgIleGlnAlaAlaArgProValArgSerValAlaLeuIleGlyP 34
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 AGAATATGCGGGTATGGCCCTTGATGAAGAATCAAGACGAGCGATGCC 150
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 heaSPfYHISglYmetLysProthrMetAlaValGlnValGlyAspArg 50
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 GTCAAAAAGGCGCAAGTCTGTTGAAGCAAAAAGNATCCGGGGGTGCT 200
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 ValIysLeuGlyGlnValLeuPheThrAspIysAsnProSerValSe 67
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
201 GTTACCGCGCGCGGTTTCAAGCAAAATCGCGCATCCATCGCGCGGAAA 250
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
67 rTyThrIaIaProGlyAlaGlyValAlaSerAlaIleHisArgGlyGlu 84
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
251 AGCGGTACTTCAGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 300
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
84 ysArgValLeuGlnSerValIleValIleAspLeuAspGlyAspGlnLeu 100
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301 GAGTTCGACGCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGANGA 350
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 GluPheAlaIaArgTyrProAlaAspIysLeuAlaThrIleuSerAlaGlu 117
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
351 ANTNNNGNCAATCTGATCCATCCATCCGTTTGGACTGCGCTGCGTANCC 400
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 nValaIArgAspAsnLeuGlnSerGlyLeuTyrThrAlaLeuArgThrA 134
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
401 GTCCGTTAGCAAAAATCCCGTCGCGCGCGCGCGCGCGCGCGCGCATCTC 450
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134 rgrProPheSerLysValProAspProGluSerProSerProSerIlePhe 150
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
451 GTCAATGCGATGACACCAATCCGTCGCGCGCGCGCGCGCGCGCGTGTGAT 500
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 ValThrAlaIleAspThrGlnProLeuAlaIleAspProGlnValValII 167
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
501 CAAGAAGACCCGCGANGATTTCAACGANGTNTCTGTGATGACCGGTT 550
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167 eAlaGlnGlnGlyGlnAlaPheGlnAlaGlyLeuThrValLeuGlyArgL 184
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
551 TGACCGAGCGTAAATCATGTGTGAAGGACGTCGGGCGCGCGCGCGCGG 600
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
184 euAla.....ArgValPheLeuGlySylsAlaGluGlyValSerLeuPro 196
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
601 TCTGAAAATGCTGCCAACATCGAAGAACATGTCGGCGCGCGCGCGATCC 650
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
199 GlyIleuAlaLeuSerGlyValThrAlaGlnAlaPheSerGlyProHisPr 215
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
651 GGCGGTTTGGTGGCGACGACATTCATTCATTCATTCATTCATTCATTCATTC 700
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
215 oAlaGlyLeuProGlyThrHisIleHisPheLeuAspProValGlyAlaG 232
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
701 ACAAAACGTTTGGACCATCAATATCAAGATCAATATTCATTCATTCATTCATTC 750
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
232 LysSerValThrPasnLeuAsnTyrGlnAspValIleAlaIleGlyLys 248
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

751 TTGTTTCGAACAGCCGCTGTGAACACCGGACCGCGGTATGCTTTGGGTGG 800
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
249 LeuPheThrThrGlyGlnLeuTyrThrGlnAlaGlyAlaIleAlaLeuAlaG1 265
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
801 TTCTCAAGTCAACAAACCGCTCTGTGACCGCTTTGGTGGCGGAAAG 850
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
265 yProValValGlnLysProAlaGlyLeuValArgThrArgLeuGlyAlaAsnL 282
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
851 TATCGCAAAATTAATCTCGCGCGGAAATTTGGTTGACGACGACAAACCGGTGAT 900
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
282 euAspGlyLeuAlaIleAlaGlyGlnLeuGlnProGlyAsnAlaGlyLeuIle 298
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
901 TCCGCTTCGCTATTCGAAGCGCGCGATTCACAAGGCGCGCGCATTAATT 950
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
299 SerGlySerValLeuGlyGlyArgThrAlaHisGlyAlaTyrAlaTyrIle 315
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
951 GGACGCTTACCAATGATTTCCGTTTGAAGAAGGCGCGCGCGGACGGAAG 1000
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
315 uGlyArgTyrHisLeuGlnLeuSerCysLeuLysGlnGlyAspGlnArg 332
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1001 AGCTGTTCGCTGGGTTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGGT 1050
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
332 LurPheLeuHisTyrLeuArgAlaGlyValGlnLysHisSerLeuLeuAsn 348
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1051 ACAGACCGTCCGCGCATTTCTTGAAAACAAATCTTCAAGTTCACGACAC 1100
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
349 ValPheValSerAlaGlyLeuGlnGlyLysArgPheAlaPheThrIleSe 365
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1101 CGTCAACGCTGGCGGACCGCGCGCGCATGTCGCGGATTCGATTCAGACGCG 1150
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
365 rThrAsnGlySerProAlaGlnAlaMetValProValGlyAsnTyrGlnAlaV 382
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1151 TAAATCGCGCTAGACATCTGCTGCTTACCTGCTTTGGCGCATTTAATGCTC 1200
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
382 alMetProLeuAspIleLeuProThrGlnLeuLeuArgTyrLeuIleVal 398
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1201 GGGGATACGACGACGCGCGCGCGCATTTGGGTTGCTTGAATTTGACGACAA 1250
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
399 GlyAspThrGlnMetAlaGlnLysLeuGlyAlaLeuGlnLeuAspIleG1 415
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1251 AGACCTCGCTTGTGCGAGCTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCG 1300
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
415 uAspLeuAlaLeuGlySerTyrValCysAlaGlyLysTyrGlnTyrGlyP 432
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1301 CGCTGTTCGCTAAGGTCTGGAACCTTGACGAGGAGCGC 1341
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
432 rolLeuLeuArgAspAsnLeuAlaArgIleGlnGlnGly 445
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

seq_name: SwissProt_40:NORA_CHLPN

seq_documentation_block:

ID NORA_CHLPN STANDARD: PRT; 467 AA.
AC 0927G2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable Na(+)-translocating NADH-quinone reductase subunit A
DE (EC 1.6.5.-) (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A)
GN (NQR complex subunit A) (NQR-1 subunit A).
DE Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales.
OX NCBI_TaxID=83556;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kaiman S., Mitchell W., Marathe R., Lamme C., Pan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";
RL Nat. Genet. 21:385-389(1999).
[2]


```

1044 CAGCGTAGACCGCTGGCCATTCTGAAACAACTCTTCAAGTTCA 1093
      |||:||||| |||:|||||:|||||
358 efnrllystrlyrleuserglyphehlyslsyrar...thrtyt 374
      |||:|||||:|||||:|||||
1094 CG.....ACAGCGTCACAGCGTGGCGCCACGTGGCCGANT 1134
      |||:|||||:|||||:|||||
374 hrasnproasphrasnleuhsiglygluttharproilletsphr 390
      |||:|||||:|||||:|||||
1135 GGTACTTACGAGCGGTAATGCGCTAGACATCTGCTTCACTCTT 1184
      |||:|||||:|||||:|||||
391 AsplletyAspllyvalmetPrometargyleProvalProleu 407
      |||:|||||:|||||:|||||
1185 GGGCGATTATCGTCGGCGATACGACGCGCCGACGATTGGTCT 1234
      |||:|||||:|||||:|||||
407 elysalavalilethrllysasnphaspleualaingluclyphel 424
      |||:|||||:|||||:|||||
1235 TGGATTGGACGACGACGACCTGCTTGTGCGCTGCTGCGCGG 1284
      |||:|||||:|||||:|||||
424 euglivalysglylualasphelaleuProthleuileasproser 440
      |||:|||||:|||||:|||||
1285 AATACGATATANGCCCGCTGTGGCTGATGCTGGAACCTTGAA 1334
      |||:|||||:|||||:|||||
441 Lysrthlglutmetleuthrllevalysgluserleuileglutylalaly 457
      |||:|||||:|||||:|||||
1335 GGAA 1338
      |||:|||||:|||||:|||||
457 SGLu 458
      |||:|||||:|||||:|||||

seq_name: SwissProt_40:NORA_CHLMU

seq_documentation_block:
ID NORA.CHLMU STANDARD; PRT; 465 AA.
AC OPRU3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable Na(+)-translocating NADH-guione reductase subunit A
DE (EC 1.6.5.-) (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A)
DE (NQR complex subunit A) (NQR-1 subunit A).
GN NORA OR TC0002.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_Taxid=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOPN / NIGG.
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Uitterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Knouri H., Craven B., Bowman C., Dodson R.,
RA Gysin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -1- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO NQR
CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC UBISEMIOUONE TO UBIQUINOL (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+) (IN) = NAD(+) +
CC UBIQUINOL + NA(+) (OUT).
CC -1- SUBUNIT: COMPOSED OF SIX SUBUNITS: NORA, NORB, NQRC, NQRD, NQRE
CC AND NQRF (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NORA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC
DR EMBL; AE002269; AAF38896.1; -
KW TIGR; TC0002; -
KW Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
KW complete proteome.
SQ SEQUENCE 465 AA; 51588 MW; 12A1AB6FA64F2C CRC64;

alignment_scores:
      Quality: 441.00      Length: 464
      Ratio: 1.521      Gaps: 14
Percent Similarity: 62.500 Percent Identity: 28.448

alignment_block:
US-09-303-518d-127 x NORA.CHLMU ..

Align seg 1/1 to: NORA.CHLMU from: 1 to: 465

4 ATTAATAATCAAAAAAGGTCTTAACCTGCGCGGCGGACGAGCGAGCA 53
|||:|||||:|||||:|||||
3 llelvalsertharglyleuaspleuserleuylsyalprolysgl 19
|||:|||||:|||||:|||||
54 AGTATTATGACGG.....CCGTCATTACCGAAGTGGCTTC 94
|||:|||||:|||||:|||||
19 UserglyphecysglylvalasproalaphelaserValaspleu. 35
|||:|||||:|||||:|||||
95 TTGCGCAAGATATGCGCGTATGCGCGCCCTNGATGAAGCAAGAAAGC 144
|||:|||||:|||||:|||||
36 ....ArgprophelalaProleupProleuglylvalysValserProgl 50
|||:|||||:|||||:|||||
145 GATCGCGTCAAAAAAGCGCAAGTGTCTTGAAGACAAAAAGNATCGGG 194
|||:|||||:|||||:|||||
51 AsplnlethrlaaglyserProleualaaglutylyserPheProgl 67
|||:|||||:|||||:|||||
195 CGTGCTTTACGCGCGCGCTTACGCAAAATCGCGCATCATCGCG 244
|||:|||||:|||||:|||||
67 yValpherlethrserservalaspllygluvalileglulearar 84
|||:|||||:|||||:|||||
245 GCGAAAGCGCGTACTTCACTGCGTGTGATGATCCGTGAA...GGCAAC 291
|||:|||||:|||||:|||||
84 lyserlysrargalaleuaseleuallevalileylslyslsProglyVal 100
|||:|||||:|||||:|||||
292 GACGAATTCAGATTGCAAGCGTACGCGCCCGAAGCGTTGGCAACTTAG 341
|||:|||||:|||||:|||||
101 SerclntrlyspheSer.....TyAspleuhsalaleuse 113
|||:|||||:|||||:|||||
342 CGGGAAGANATNGNNGCAATGATCAATCCAGTGTGACACGCGC 391
|||:|||||:|||||:|||||
113 rGlnlysgluleuengluvalPheylslysglygluleuphetrleup 130
|||:|||||:|||||:|||||
392 TGGGTATNCCTCGCTCAGCAAAATCCGTGCGTGCATGCGGAG...CCG 438
|||:|||||:|||||:|||||
130 helysglnArgprophelasn...lleProalaleuProthlrglnserPro 145
|||:|||||:|||||:|||||
439 TTGGCATCTTCTGCAATGCGATGACACCAATCGCTN.....GCGGC 482
|||:|||||:|||||:|||||
146 ArgasvalaphelleanleualaaspasnargProphenrProserVa 162
|||:|||||:|||||:|||||
483 AGACCCGTGTGTGTATCAAAAGACCCGCGANAT.....T 520
|||:|||||:|||||:|||||
162 lGlnlyshlsleuserleupheserlysgluasgllytyrtyllep 179
|||:|||||:|||||:|||||
521 TCAGAGGANGTNGCTGATATGAGCGGTTCAGCGGCTAAATCAAT 570
|||:|||||:|||||:|||||
179 hevalValglValalGlnalalelalsleuphelyleuysProhls 195
|||:|||||:|||||:|||||
571 GTGTGTAAGGCACTGGCGACAGCTGCGTGTGAATGTGCGCAACAT 620
|||:|||||:|||||:|||||
196 llevalserthrasparleuserleuProthlrglnaspleuileserVa 212
|||:|||||:|||||:|||||
621 C...GAACACATGATTCGGCGCGCGCATCCGCGGCTTTGAGGCGA 667
|||:|||||:|||||:|||||
212 lAlahlsleuhsrthrllealalelyProtyrProserlyserProserT 229
|||:|||||:|||||:|||||

```



```

115 saspleuLeuGluValPheLysLysGluGlyLeuPheAlaLeuPheLysG 132
132 InArgProPheAsp...IleProAlaLeuProThrGlnSerProHisIleIle 147
445 ATCTTGCTCAATGCGATGACACCAATCCGCTN.....GCGGACAGACC 488
148 ValPheIleAsnLeuAlaAspAsnArgProPheThrProSerValGluIly 164
489 TGTGGTTGTCATCAAGAAGCCGNCANGAT.....TTCAGAC 526
164 shLseuSerLeuPheSerSerLysGluAspGlyTyrTyrIlePheVal 181
527 GANGTWTGCTGATGAGCGGTTTGACCGAGCGTAAATTCATGCTGT 576
181 alGlyAlaGlnAlaIleAlaLysLeuPheGlyLeuLysProHisIleIle 197
577 AAGGACGCTGGCGCAGACGTCGCTGAAAATGCTGCCACATC...GA 623
198 SerThrAspArgLeuThrLeuProThrGlnAspLeuValSerIleAlaH 214
624 AACACATGATTCGGCGCCGCGATCCGCGGTTGATGCGACGACCA 673
214 sleuHisThrIleAspGlyProPheProSerGlySerProSerThrHisI 231
674 TTCATTTTCATGAGCCGTC...GGTCAACAAAACCGTTGGAGCCATC 720
231 leHisHisIleAlaArgIleAlaGlnAspGlnArgAspAlaValPheThrIle 247
721 AATTATCAAGATGTAATTCATCGACGCTGTTGTTGCAACGCGCGCT 770
248 SerPheGlnGluValLeuSerIleGlyHisLeuPheLeuLysGlyPhe 264
771 GAACACGAGCGGTGATGCTTGGGTGCTTCTCAATC.....AAC 814
264 lleuGlyGlnGluIleValAlaLeuAlaGlySerAlaLeuProProSerG 281
815 AACACGCGCTTGGCTGACGTTTGGCGAAGTATCGCAATTAAT 864
281 InArgLysIleuThrIleThrAlaLysGlyAlaSerPheSerPheLeu 297
865 GCGGCGCAATGGTGGTACGACGACCAAC...CGCGTATTCGCGTTCGT 911
298 ProLysAspIlePheSerSerAspGlnIleThrLeuIleSerGlyAsp 314
912 ATTGAACGCGCGGATTTACACAGCGCGACAT...TATTGGGACGT 958
314 oleuThrGlyArgLeuLysLysGluGlnAsnProCysLeuGlyMet 331
959 ACCACATCAGATTCCTGATTCGAAGAAGCGCGACGAGAGCGTGT 1008
331 rGAspHisThrIleThrLeuLeuProAsnProLysThrArgLysSerPhe 347
1009 GCGTGGGTGGCGCGACGCGACCAATATCCATACGCGGTACGACCT 1058
348 SerPheLeuArgLeuGlyTyrAsnLysLeuThrValThrArgThrTyle 364
1059 CGGCGATTCCTGAAAACAAA.....CTCTCAAGTTTCAGACGACG 1102
364 uSerGlyPhePheLysArgLysArgValPheMetAspMetAspThrAsn 381
1103 TCACGCGTGGCGACCGCGCATGTGCGATGTGACTTACGAGCGCTA 1152
381 eHisGlyGlyLysArgProIleIleAspAlaGlnIleTyrGlnArgVal 397
1153 ATGCGCGTACAGATCTGCTACCTGCTTTGCGCGATTAATTCGCGG 1202
398 SerAlaIleProValProValAlaLeuIleIleLysAlaLeuGluThrG 414
1203 CGATACGCGACGCGCAAGCATTTGGTGTGATTTGACGAGAGAG 1252

```

```

414 nasnPhelGluAlaCysArgLeuGlyLeuLeuGluValAlaProGlu 431
1253 ACCTGCTTTGGCAGCTTCGTCCTGCGCCGCAATACGAATANGCCG 1302
431 sPhelAlaLeuProThrPheIleAspProSerLysThrGluMetPheSer 447
1303 CTCTGCGTAAGTGTGTAACGNTTGAGAGAA 1338
448 lleValLysGluSerLeuLeuArgTyrAlaLysGlu 459

seq_name: SwissProt_40:RNFC_VIBCH
seq_documentation_block:
ID RNFC_VIBCH STANDARD; PRT; 774 AA.
AC O9R88;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE Electron transport complex protein rnfC.
GN RNFC OR VC1015
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_taxid=66;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Olin H., Dragoi I., Sellers P.,
RA McDonald L., Ueberlack T., Fleischmann R.D., Niemann W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
CC -1- FUNCTION: May be part of a membrane complex involved in electron
CC transport (By similarity).
CC -1- CORPORA: Binds 2 4Fe-4S clusters (Potential).
CC -1- SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
CC rnfD, rnfE and rnfG (By similarity).
CC -1- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -1- SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC RNFC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AEO04183; AAF94176.1; ALT_INT.
DR TIGR; VC1015;
DR InterPro: IPR001450; 4Fe4S_ferredoxin.
DR InterPro: IPR001949; Complex1_51K.
DR Pfam; PF01512; Complex1_51K; 1.
DR Pfam; PF00037; fer4; 2.
DR PROSITE; PS00198; 4Fe4S_FERREDOXIN; 2.
KW Electron transport; Iron-sulfur; 4Fe-4S; Inner membrane;
KW Complete proteome.
FT METAL 378
FT METAL 381
FT METAL 384
FT METAL 388
FT METAL 417
FT METAL 420
FT METAL 423
FT METAL 427
FT METAL 427
FT SEQUENCE 774 AA; 83500 MM; 44B1E3C304E3AE CRC64;

```

alignment_scores:

Quality: 157.00 Length: 472
 Ratio: 0.720 Gaps: 18
 Percent similarity: 46.186 Percent identity: 21.822

alignment_block:

US-09-303-518D-127 x RNFC_VIBCH ..

Align seg 1/1 to: RNFC_VIBCH from: 1 to: 774

```

52 CAAGTCATTATGACGGCGCCGTCATTACGAAATCGCGTTGCTGGCGA 101
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : : :
32 GlnProIleIleAsnAlaSerIleProAsnGlnLeuValLeuProLeu 48
   : : : : : : : : : : : : : : : : : : : : : : : :
102 ACAATATGCGCGTATGCGCCCTNGATGAAATGCAAGAGCGCATGCCG 151
   : : : : : : : : : : : : : : : : : : : : : : : :
48 sGlnHisIleGlyLysAlaGlyAspLeuLeuValLysValGlyAspArg 65
   : : : : : : : : : : : : : : : : : : : : : : : :
152 TCAAAAAGGCGCAAGTGTCTTTGAAGACAAAAGNATCCGCGCGTGTG 201
   || ||||| ||| : : : : : : : : : : : : : : : : : :
65 alLeuLysGlnProLeuThrGlnTyrThrSerThrPheMetLeuPro 81
   : : : : : : : : : : : : : : : : : : : : : : : :
202 TTTACCGCGCCGTTTACGCAAAATCGCGCCATCCATCGCGCGAATA 251
   : : : : : : : : : : : : : : : : : : : : : : : :
82 IleHisIleProThrSerGlyValIleSerAlaIle.....GluPr 95
   : : : : : : : : : : : : : : : : : : : : : : : :
252 GCGCGACTTCAG.....TCGCGTGTGATTG 277
   ||||| : : : : : : : : : : : : : : : : : : : : : :
95 oArgThrValAlaHisProSerGlyLeuSerGlyLeuCysIleValLeu 112
   : : : : : : : : : : : : : : : : : : : : : : : :
278 CCGTTGAAGCAACGAGAA.....ATCGAGTTC 306
   : : : : : : : : : : : : : : : : : : : : : : : :
112 hrProAspGlnGlnGluTyrPheGlnLeuGlnProGlnProAspPhe 128
   : : : : : : : : : : : : : : : : : : : : : : : :
307 GAACGCTACGCGCCGAGACGTTGGCAACTTAAGCGCGGANGAANTNG 356
   : : : : : : : : : : : : : : : : : : : : : : : :
129 GlnGlnLeuThrProGlnThrLeuLeuGlnLeuIleArgGlnAlaGly 145
   : : : : : : : : : : : : : : : : : : : : : : : :
357 NNGCAATCGATCCAAATCGGTTTGAGCTGCGTGGCTANCGTCGCT 406
   : : : : : : : : : : : : : : : : : : : : : : : :
145 eSerGlyMetGlyGlyAlaGlyPheProThra..... 156
   : : : : : : : : : : : : : : : : : : : : : : : :
407 TCAGCAAAATCCCTGCGCTGATGCGAGCCGTCGCCATCTTCGTCAAT 456
   : : : : : : : : : : : : : : : : : : : : : : : :
157 ..LysLysLeuGlnSerGlyLeuSerArgThrGlnIleLeuIleLeasn 172
   : : : : : : : : : : : : : : : : : : : : : : : :
457 GCGATGACACCAATCCG...CTNGCGGACGACCTGTGTGTGTATCAA 503
   ||| : : : : : : : : : : : : : : : : : : : : : :
173 AlaAlaGlnCysGlnProTyrIleThrAlaAspValLeuMetArgGln 189
   : : : : : : : : : : : : : : : : : : : : : : : :
504 AGAAGCCGCGANGATTCAGACGANGTNGCTGGA..... 540
   : : : : : : : : : : : : : : : : : : : : : : : :
189 nTyrAlaHisGlnIleIleGlnGlyIleGlnIleValGlnHisIleLeu 206
   : : : : : : : : : : : : : : : : : : : : : : : :
540 ..... 540
206 ysrProLysLeuThrIleIleGlyIleGlnAspAsnLysProGlnAlaVal 222
   : : : : : : : : : : : : : : : : : : : : : : : :
541 .....TTGAGCCGTTTGACCGACGCTAAATCCATGTGTGAAGCGAC 584
   ||| : : : : : : : : : : : : : : : : : : : : : :
223 AlaAlaLeuGlnGlnAlaAlaGlnAspLysPrometValIleArgValI 239
   : : : : : : : : : : : : : : : : : : : : : : : :
585 TGCGCGAGACGTGCGCT.....GAAATGTGCTG 613
   : : : : : : : : : : : : : : : : : : : : : : : :
239 eProThrLysTyrProSerGlyGlyLysGlnLeuIleLysIleLeu 256
   : : : : : : : : : : : : : : : : : : : : : : : :
614 CCAACATCGAACAACATGAATTCGGCGCGCGCATCCGCGGTTTGACT 663
   : : : : : : : : : : : : : : : : : : : : : : : :
256 hrAsnLeuGlnValProLysGlyGlyIleProAlaAspIleGlyLeu... 271
   : : : : : : : : : : : : : : : : : : : : : : : :
664 GGCACGCAATTCATTGACGCGGTGCGTGCACAAACAAACCGTTTG 713
   : : : : : : : : : : : : : : : : : : : : : : : :

```

```

272 .....MetValGlnAsnValGlySer..... 278
714 GACCATCATATTATCAGATGTAAATTCGATCCGACGCTTTGTTGCAACG 763
   : : : : : : : : : : : : : : : : : : : : : : : :
279 .....LeuGlnIleAlaIleAlaGlyAlaIleValHisG 289
   : : : : : : : : : : : : : : : : : : : : : : : :
764 GCCGCTGGAACACCGAGCGCGTATGCTTTGGGTGTTCTCAAGTCAAC 813
   : : : : : : : : : : : : : : : : : : : : : : : :
289 LysIleProLeuIleArgArgValValThrLeuThrLysAspCysPheArg 305
   : : : : : : : : : : : : : : : : : : : : : : : :
814 AAACACAGCCTCTGGGTACACGCTTTGGGTGCGCAAGTATCCGAATTAC 863
   : : : : : : : : : : : : : : : : : : : : : : : :
306 LysProArgAsnValIleProAlaLeuLeuGlyThrProValGlnAlaLeu 322
   : : : : : : : : : : : : : : : : : : : : : : : :
864 TGCGGCGCAATGCTGTGACGACGACAAC.....CGCGTATTTCCG 904
   : : : : : : : : : : : : : : : : : : : : : : : :
322 uAsnGlnPheGlyTyrLysAlaAspLysLysLeuProAlaGlyIleMetG 339
   : : : : : : : : : : : : : : : : : : : : : : : :
905 GTTCGCTATTGAACGCGCGATTCACACAGCGCGCACGATTATTGGGA 954
   : : : : : : : : : : : : : : : : : : : : : : : :
339 LysIleProMetMetGlyPheThrLeuPro..... 348
   : : : : : : : : : : : : : : : : : : : : : : : :
955 CGCTACACAAATGAGATTCGCTATTCGAAGAAGCGCGCAAGAGCT 1004
   : : : : : : : : : : : : : : : : : : : : : : : :
349 .....HisAlaGlnValProIle.....ThrLysThrAl 358
   : : : : : : : : : : : : : : : : : : : : : : : :
1005 GTTCGCGTGGGTTGCGCGCGACCGGACAAATATCATCATCCGCTAGCA 1054
   : : : : : : : : : : : : : : : : : : : : : : : :
358 aaAsnCysIleLeuAlaPro.....ThrArgAsnG 368
   : : : : : : : : : : : : : : : : : : : : : : : :
1055 CCGTCGCGCAATTTCGTGAATAACAACTTTCAAGTTACAGACGCCGTC 1104
   : : : : : : : : : : : : : : : : : : : : : : : :
368 LysLeu.....ThrSerSerAsp 373
   : : : : : : : : : : : : : : : : : : : : : : : :
1105 AACGCTGCGCGCGCCATGTCGCGATGTGTAATTACGAGCGCTAAT 1154
   : : : : : : : : : : : : : : : : : : : : : : : :
374 Asn.....GlnMetAlaCysIleArgCysGlyGlnCysAlaGlnLac 388
   : : : : : : : : : : : : : : : : : : : : : : : :
1155 GCCGCTGAGACATCCGCTACCCCTGTTTGGCGATTAATGTCGCGCG 1204
   : : : : : : : : : : : : : : : : : : : : : : : :
388 sProValSerLeuLeuProGlnGlnLeuGlnTyrHisAlaLysAlaGln 405
   : : : : : : : : : : : : : : : : : : : : : : : :
1205 ATACGACACGCGCGCAAGCATTTGGTTCCTTGAATGCGAAGAAGAC 1254
   : : : : : : : : : : : : : : : : : : : : : : : :
405 LysPheAspLys.....CysGlnGlnLeuAspLeuLysAsp 416
   : : : : : : : : : : : : : : : : : : : : : : : :
1255 .....CTCGCTTGTGCAGCTTCGTGCGCGCGCAATACAGATA 1295
   : : : : : : : : : : : : : : : : : : : : : : : :
417 CysIleIleLysGlyAlaCysAlaTyrValCysProSerGlnIle..... 431
   : : : : : : : : : : : : : : : : : : : : : : : :
1296 NGCGCCGCTGTGCGT 1311
   : : : : : : : : : : : : : : : : : : : : : : : :
432 .....ProLeuValGln 435
   : : : : : : : : : : : : : : : : : : : : : : : :

```

seq_name: SwissProt_40:HKR1_YEAST

seq_documentation_block:

ID HKR1_YEAST STANDARD; PRT; 1802 AA.
 AC P41809;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Hansenula WRAKII killer toxin-resistant protein 1 precursor.
 GN HKR1 OR YDR420W.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YNN 295;
 RX MEDLINE=94156857; PubMed=8113191;
 RA Kasahara S., Yamada H., Mio T., Shitatori Y., Miyamoto C.,


```

435 CTCGGCATCGACGAGGATTTGTGTAACGAGCGTACGCGCAG 386
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
647 gserleuserthrser.....AsnAlaThrSerAlaSerAsnV 660
385 TCACAAACCGGATTCAGATTCGNNNNANTTCGCGCGTTAAG 336
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
660 AlProThrGlyThrPheSerSerMetSerSerHisThrSerValLehr 676
335 .....TTGGCAACGCTTCGGCGCGTAGCGTTGAACCTGATT..... 297
677 ProGlyPheSerThrSerSerAlaSerLeuAlaIleAsnSerThrValVa 693
296 ....TCGTCGTCCTTCACGCAATCAGCAGCTGAGTACGCGCT 251
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
693 lserSerSerLeuAlaGlyTyrSerPheSerThrProGlySerProT 710
250 TTTCGCGCGGATGATGCGCGGATTTGCTGAAACGCGCGTAAAC 201
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
710 hrThrSerThrLeuValThrSerGluAlaProSerThrValSerSerMet 726
200 ACCACGCGCGATNCTTTTGCT.....TCAAA 172
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
727 ThrThrSerAlaProPheIleAsnSerThrSerAlaAlaGProSerPr 743
171 CAGCAGCTGGCGCTTTTTCAGCGCATCGCTTCC..... 138
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
743 oSerThrAlaSerPheIleThrGlySerThrSerSerLeSerSerValP 760
137 .....TTGACTTTTCATCMAGGCGCATACCG 111
760 roLeuAlaSerGlyAspValThrSerSerSerLeuAlaAlaHisAsnLeuThr 776
110 GCATATTCCTGCCAAGCAGCAGCTCGGTAATGAGCGCGCGTCA 63
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
777 ThrPheSerAlaProSerThrSerSerAlaGlnLeuValSerLysSer 792

```

seq.name: SwissProt_40:NI21_RAT

seq_documentation_block:

```

ID NI21_RAT STANDARD; PRT; 1199 AA.
AC P52591;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Nuclear envelope pore membrane protein POM 121 (Pore membrane protein
of 121 kDa) (P145).
GN NUP121.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RX MEDLINE=93328754; PubMed=8335683;
RA Hallberg E., Wozniak R.W., Blobel G.;
RT "An integral membrane protein of the pore membrane domain of the
nuclear envelope contains a nucleoporin-like region."
RL J. Cell Biol. 122:513-521(1993).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF THE NUCLEAR PORE COMPLEX. THE
REPEAT-CONTAINING DOMAIN MAY BE INVOLVED IN ANCHORING COMPONENTS
OF THE PORE COMPLEX TO THE PORE MEMBRANE.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. NUCLEAR PORE
MEMBRANE.
CC -1- DOMAIN: CONTAINS X-F-X-F-G REPEATS.
CC -1- SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS
SEPARATED BY SER/THR-RICH DOMAINS IS SIMILAR TO THAT OF YEAST
NUP1, NSP1 AND MAMMALIAN P62 AND NUP153.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

DR EMBL: Z21513; CAA79275.1; -
DR EMBL: Z21514; CAA79276.1; -
KW Nuclear protein; Transport; Transmembrane; Repeat.
FT DOMAIN 1 56 CISTERNALE SIDE (POTENTIAL).
FT DOMAIN 57 77 POTENTIAL.
FT DOMAIN 76 1199 PORE SIDE (POTENTIAL).
FT DOMAIN 4 8 POLY-ALA.
FT DOMAIN 53 58 POLY-ALA.
FT DOMAIN 438 441 POLY-SER.
FT DOMAIN 497 500 POLY-PRO.
FT DOMAIN 684 687 POLY-SER.
FT DOMAIN 1023 1026 POLY-ALA.
FT DOMAIN 1033 1038 POLY-SER.
FT DOMAIN 1058 1063 POLY-GLY.
SO SEQUENCE 1199 AA; 120784 MW; 6DC4451B91D5B907 CRC64;

```

alignment_scores: Quality: 123.50 Length: 468
Ratio: 0.542 Gaps: 17
Percent Similarity: 48.718 Percent Identity: 20.513

alignment_block:

US-09-303-518D-127/rev x NI21_RAT ..

Align seg 1/1 to: NI21_RAT from: 1 to: 1199

```

1340 CCTTCCTTCACANGTTTCACACCTTACGACACGCGCGCNTATTC 1291
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
546 ProPheThrLeuThrLeuProThrValGlyProAlaAlaSerProAl 562
1290 GATTTGCGCGCGCAGACG.....AAGCTGCACAAAG 1259
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
562 AserLeuProAlaProSerSerAsnProLeuLeuGlySerLeuLysW 579
1258 CGAGTCTTCT...TCGTCATTCACGACCAACCAATCTTCGCGCTG 1212
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
579 etGlnLysSerProAlaProSerSerSerGlnProGlyAlaAlaThr 595
1211 TCGGTATCCGCCAGCTTAATCGCCCAACGAGGTGCGCAGATGTC 1162
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
596 ValAlaAlaProSerProPylsThrProSerLeuLeuAlaProLeuVa 612
1161 TAGCGGATTTACGCGCTCGTAAGTACCAATC...GGCACCATGGCGGCT 1115
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
612 lserProLeuThrGly.....ProLeuAlaSerThrSerSerAspS 626
1114 CGCCACCGTTGACGCGCTGTGTAACCTGAGAGT..... 1080
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
626 erLysProThrThrThrPheLeuGlyLeuAlaSerAlaSerSerAlaThr 642
1079 ...TTGTTTTCAGGAATGCGCGAGCGTCGTACCGTGATGAGTATT 1033
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
643 ProLeuThrAspThrLysAlaProGlyValSerGlnAlaGlnLeuCySva 659
1032 GTTCGCGCTCGCGCAACCGCAGACAGCTCT..... 999
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
659 lserThrProAlaAlaThrAlaProSerProThrProAlaSerThrLeuP 676
998 .....TTGCTGCGGCGCTTTCGATACGGAATCGATTGTGGTAG 957
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
676 hecLysMetLeuSerProProAlaSerSerSerSerLeu.....Ala 689
956 CGGCCAATATATCGGCGCGCGCTGTGTAATCGCGCGCTTC..... 915
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
690 ThrProGlyProAlaCysAlaSerProMetPheLysProIlePheProAl 706
914 .....AATACGACCGGAATATCAGCGGTTGTCTCGCT 881

```

```

706 aThrProlysergluserasphnProLeuProThrSerSerSerAla 723
      ::::: ||| |||||
880 CAACCAATTCGCCGCGAGTAATTTGGCATCTTTCGACCCAAACGTA 831
      ::::: ||| |||||
723 laThrThrThrProAla.....SerThrAlaLeuProThrThrAla 736
      ::::: ||| |||||
830 CGCAAGAGCGCTGTTGTTACTTGAGAACCCACCAAGCAATCAGCGG 781
      ::::: ||| |||||
737 .....ThrAlaThrAla 740
      ::::: ||| |||||
780 CTCGGTGTTCAGACGGCCTTTGCAACCAACGTCGATGGCAATTCAT 731
      ::::: ||| |||||
740 aHisThrPheLysProIlePhegluserValGluProPheAla..... 754
      ::::: ||| |||||
730 CTGTATTAATGATGGTCCAAACGTTTGTTCAGACGCGCTCATATG 681
      ::::: ||| |||||
755 .....AlaMetProLeuSerProProPheSerLeu 764
      ::::: ||| |||||
680 AATGCAATGCGTCCGCACTCAACCGCGCGATGGCGCGCGCAATTC 631
      ::::: ||| |||||
765 LysGlnThrThrAlaProIleThrThrAlaThrSerAlaPro..... 779
      ::::: ||| |||||
630 ATGTGTTTCAGATGTCGACATTTTCAGACGACGTCGCGCAGCTG 581
      ::::: ||| |||||
780 .....LeuLeuThrGlyLeuGlyThrAlaThrSerThrValAla 793
      ::::: ||| |||||
580 CTTACACACATGATTTTACGCTCGTCAACGCGCTCAATACACACANA 531
      ::::: ||| |||||
793 hr.....GlyThrThrAla 797
      ::::: ||| |||||
530 CMTGCTGAATCMTGCGGCTTTGATCAACACACAGGAGGTGCG 481
      ::::: ||| |||||
798 SerAlaSerLysProValPheGlyPheGlyValThrThrAlaAlaSerTh 814
      ::::: ||| |||||
480 CGCNAAGCGATGTCATCCATCGCATTCAGACAGATGGCGAGCTCGG 431
      ::::: ||| |||||
814 rAlaSerThrIleAlaSerThrSerGlnSerIleLeuPheGlyGlyAla 831
      ::::: ||| |||||
430 CATCAACGCGAGGATTTTGGCTGAACGACGCGATTCAGACGCGCATC 381
      ::::: ||| |||||
831 rProValThr.....AlaSerSerSer 838
      ::::: ||| |||||
380 AAACCGGATGATGATGATGCGCNCNNMNTTCGCGCTT...AAGTT 334
      ::::: ||| |||||
833 AlaProAlaLeuAlaSerIlePheGlnPheGlyLysProLeuAlaProAl 855
      ::::: ||| |||||
333 TGCCAAAGCTTGGCGCGGTAGCGTTGCAACTGATTTGCTGTCCTT 284
      ::::: ||| |||||
855 aAlaSerValAlaGlyThrSerPheSerGlnSerIleAlaSerAlaG 872
      ::::: ||| |||||
283 CAACGCGCATTCAGACGCGCTGATTCGCGCTTTCCGCGGATGATG 234
      ::::: ||| |||||
872 lThrAlaAlaSerAsn..... 877
      ::::: ||| |||||
233 GCGGCGATTTGCTGAACNGCGCGGTAAACACACGCGCGATTCCT 184
      ::::: ||| |||||
878 .....SerSerIleGlyPheSerGlyPheGlyGlyThrLe 889
      ::::: ||| |||||
183 TTTGTCCTCAACAGACTTGGCTTTTTCAGCGCATCGCTTCTCTGA 134
      ::::: ||| |||||
889 uThrThrSerThrSerAla...ProAlaThrThrSerGlnProThrLeu 905
      ::::: ||| |||||
133 CTTTATCAAGGGGCGCATACCGGATTCCTTCGCAAGACGCGACT 84
      ::::: ||| |||||
905 hrPheSerAsnThrValThrProThrPheAsnIleProPheSerAlaSer 921
      ::::: ||| |||||
83 TCGGTAAATGACGGCGCTCATAAATGACTTCTCGGTCGCGCGAT 34
      ::::: ||| |||||
922 AlaLysProAlaLeuProThrThrTyrrProGlyAlaAsnSerGlnProThrPh 938
      ::::: ||| |||||
33 GGGC 30
      ::::: ||| |||||

```

```

938 ecly 939
seq_name: SwissProt_40:AMYH_YEAST
seq_id: AMYH_YEAST
seq_block:
ID AMYH_YEAST STANDARD; PRT; 1367 AA.
AC P08640; P08068;
DT 01-AUG-1988 (Rel. 08, Created)
DR 01-FEB-1995 (Rel. 31, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glucoamylase S1/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
DE glucosidase) (1,4-alpha-D-glucan glucohydrolase).
GN STR1 OR STR2 OR MAUS OR Y1R019C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID:4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Barrett B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagers K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RA Submitted (Dec-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX MEDLINE=87194600; Pubmed=3106330;
RT Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
RT STR1."
RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=SPX101-1C;
RX MEDLINE=89031230; Pubmed=3141213;
RT Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STR2 and SGA genes
RT from Saccharomyces cerevisiae."
RL FEBS Lett. 239:179-184(1988).
CC -i- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -i- SIMILARITY: TO S. POMBE SPC215.13.
CC -i- SIMILARITY: SOME, TO S. POMBE SPC285.13C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; Z38061; CAA86176.1; -
CC DR EMBL; M16164; AAA35014.1; -
CC DR EMBL; M16165; AAA35015.1; -
CC DR EMBL; X13857; CAA32069.1; -
CC DR PIR; B26877; B26877.
CC DR PIR; A26877; A26877.
CC DR PIR; S48478; S48478.
CC DR SGD; S0001458; M0C1.
CC KM Hydrolyase; Glycosidase; Polysaccharide degradation; glycoprotein;
CC signal; Multigene family.
CC KW SIGNAL;
CC FT CHAIN 1 21
CC FT CHAIN 22 1367
CC FT DOMAIN 210 1367
CC FT CARBOHYD 817 817
CC FT CARBOHYD 874 874
CC FT SEQUENCE 1367 AA; 136110 MW; 91C00E2DBD61AA9D CRC64;

```


alignment_scores:
 Quality: 123.00 Length: 532
 Ratio: 0.498 Gaps: 23
 Percent Similarity: 46.429 Percent Identity: 20.489

alignment block:

US-09-303-518d-127/rev x AMYH_YEAST ..

Align seg 1/1 to: AMYH_YEAST from: 1 to: 1367

```

1337 TCCTCTCTCAANGTTTCACGACCTTAGCAACAGCGGCGCCATTATCGTA 1288
|||||
891 SerPheSerThrGlyThrThrValThrProSerSerLysTyr.... 905
1287 TTTGCCGCGGACGACGAGCTGCACAAAGCAGTCTTCTTCGTCACATT 1238
|||||
906 ...ProGlySerGlnThrGluThrSerValSerSerThrThrGluThr 921
1237 CCACGACCAACCAATGCTGCGCGCTGCGTA.....TCGCGG 1200
|||||
921 ThrLeuAlaProThrLysThrThrThrSerValThrThrProSerThrThr 937
1199 ACGATTAAATGCGCGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1150
|||||
938 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 954
1149 GCGCTCGTAAGTACCAATCGCACCAGGCGCGCGCGCGCGCGCGCGCGG 1100
|||||
954 ThrSerGlyCysSerProLysThrValThrThrThrValProThrThrThr 971
1099 CTGTGCTGACCTGGAAGTTGTTTTCAGGAATGCGCGAGGCGTGTGTA 1050
|||||
971 ThrThrThrSerValThrThr.....SerSerThrThrThrThrThr 983
1049 GCGGTGATGAGATATTTGCTCCGCTGCGCGCGCAACCGACGAGCAGCTC 1000
|||||
984 ThrThrThrValCysSerThrGlyThrAsnSerAlaGlyGluThrThrSe 1000
999 TTTGCTGCGCGCTTCTTGATTAACGGAATCTGATGTAAGCGTCCCA 950
|||||
1000 rGlyCysSerProLysThrThr..... 1008
949 AATATGCTGCGCGCTGCTGTATTCGCGCGCGCTCATACGCAACCGGAA 900
|||||
1009 ...ThrThrValProCysSerThrSerProSerGluThrAlaSerGlu 1023
900 ..... 900
1024 SerThrThrThrSerProThrThrProValThrThrValAlaSerThrThr 1040
899 .....ATCAGCGCGTGTGCTGCTGCAACCAATTCGCCGCGAGTAATT 857
|||||
1040 rValValThrThrGluThrLysSerThrThrLysProGlyGlyGluThr 1057
856 GCGATTAATCTTGCACCAAAAGGCTAGCAAGAGCGCTGCTTCTTACT 807
|||||
1057 ThrThrThrPheValThrLysAsnLeu...ProThrThrThrLysLeuThrThr 1072
806 TGAGAACACCAAGCAATCAGCGCGCTGCTGCTTACAGAGCGCTGTGCG 757
|||||
1073 ThrLeuAlaProThrProSerValThr.....ThrValThr 1083
756 AAACAACAACGTCGATGCGCATTTACATTTGATTAATGATGTCGCAACGG 707
|||||
1083 rAsnPheThrProThrThrThrThrThr.....ThrVal 1094
706 TTTTGTTCACGACGCGCGCATGAATGAATGAATGTCGCGCACATCAAA 657
|||||
1094 aLysSerThrGlyThrAsnSer.....AlaGlyGluThr 1105
656 CCGGCGGATGCGGCGCGCAATTCATGTTGATGATGTTGCGACGAT 607
|||||

```

```

1106 ThrSerGlyCysSerProLysThrValThrThrThrValProCysSerThr 1122
606 TTCAGAGCGC.....ACGCTGCGCGACGCTGCTTACACACATGCA 566
|||||
1122 rGlyThrGlyGluThrThrThrGluAlaThrThrLeuValThrThrAlaVal 1139
565 TTTTACGCTCGTCGCAACGCGCTCAATACAGACANACNTGCTGAATCA 516
|||||
1139 AlThrThrThrValThrThrThrGluSerSerThr.....Gly 1151
515 TCGACGCGCTTCTTGATCAACACAGAGGCTGCGCGCAACGCGATTTGT 466
|||||
1152 ThrAsnSerAlaGlyThrThrThrThrThrGlyTyrThrThrLysSerValPr 1168
465 GTCCATGCGATTGACGAGATGCGAAGCGCTGCGCATGACGCGCAGGGA 416
|||||
1168 oThrThrThrValThrThrThrLeuAlaProSerAlaProValThrPro... 1183
415 TTTTGTGAACGAGGAGGNTACGACGCGCAGTCCACAAACCGGATTGGATC 366
|||||
1184 .....AlaThrAsnAlaVal.....ProThrThrThr 1192
365 AGATTGCGNNMNTNTGCGCGCTTACGTTGCCAGCGCTTGGCGCGC 316
|||||
1193 ThrThrThrGluCysSerAlaAla.....ThrAsnAlaAlaGly.. 1205
315 GTAGCGTTCGACCTGATTCGTCG..... 291
1206 ...GluThrThrSerValCysSerAlaLysThrThrLeuAlaSerSerAla 1221
291 ..... 291
1221 eraLaglyGluAsnThrAlaProSerAlaThrThrProValThrThrAla 1237
290 TTGCTTCACGCGCATGACGACGACGACGACGACGACGACGACGACGACG 258
|||||
1238 ThrProThrThrValThrThrThrThrThrThrThrThrThrThrThrThr 1254
257 .....ACGCGCTTTCGCGG...CGATGGA 236
1254 aGlyGluThrThrThrGlyTyrThrThrThrLysSerThrProThrThrTyr 1271
235 TGCGCGCGATTTTGCT.....GAAACGCGC 210
|||||
1271 ThrThrThrLeuThrProGlySerAsnGlyAlaLysAsnTyrGluThrVal 1287
209 GCGGTAAACACGCGCGC.....GGATN 187
1288 AlaThrAlaThrAsnProLysSerThrThrThrSerGlnLeuAlaThr 1304
186 CTTTGTGCTTCACAGCAGCTGCGCTTTTGAAGCGATGCGCTTCT 137
|||||
1304 ThrThrAlaThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1320
136 TGACTTTCATCNAGGGCGCATACGCGCATATTCCTGCGCAACGAGCG 87
|||||
1320 ThrThrGly.....ProLeuGlnSerAlaSerGlySerAla 1331
86 ACTTGTGATGACGCGCGCGCATTAATGACT.....TCGTC 49
1332 ValAlaThrThrSerValProSerThrThrThrThrThrThrThrThrThr 1348
48 CCGTCTGCGCGGATGCGGAGGCTTGAAGCTTTTGAATTTTAATC 3
|||||
1348 AsnThrLeuValThrGlyAsnPheMetThrPheLeuLeuAlaLeu 1363

```

seq_name: SwissProt.40:RNFC_PSEAE

seq_documentation_block:

ID RNFC_PSEAE STANDARD; PRT; 774 AA.

AC O9HYB8;

DT 01-MAR-2002 (Rel. 41, Created)

DT 01-MAR-2002 (Rel. 41, Last sequence update)

[illegible]

```

seq_documentation_block:
ID      VGLX_HSVEB          STANDARD;          PRT;          797 AA.
AC      P28968:
DT      01-DEC-1992 (Rel. 24, Created)
DT      01-DEC-1992 (Rel. 24, Last sequence update)
DE      01-DEC-1992 (Rel. 24, Last annotation update)
DE      Glycoprotein X precursor.
GN      71.
OS      Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC      Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC      Alphaherpesvirinae; Varicelloviruses.
OX      NCBI_TaxID=31520;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=92295566; PubMed=1318606;
RA      Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RL      "The DNA sequence of equine herpesvirus-1";
        Virology 189:304-316 (1992).
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/isb-sib.ch).
CC      or send an email to license@isb-sib.ch).
-----
DR      EMBL; M86664; AAB02506.1; -.
DR      PIR; H36802; VGBEX1.
DR      Glycoprotein; Transmembrane; Signal.
PT      SIGNAL
                1              22

```

```

FT CHAIN 23 797 GLYCOPROTEIN X.
FT DOMAIN 23 465 SER/THR-RICH.
FT TRANSMEM 766 790 POTENTIAL..
FT CARBOHYD 590 590 N-LINKED (GLCNAC....) (POTENTIAL)
SQ SEQUENCE 797 AA; 80342 MW; 50C9ED9211F5E5B2 CRC64;

alignment_scores:
    Quality: 116.50      Length: 450
    Ratio: 0.565        Gaps: 15
Percent Similarity: 45.778    Percent Identity: 21.111

alignment_block:
US-09-303-518D-127/rev x VGLX_HSVEB ..

Align seg 1/1 to: VGLX_HSVEB from: 1 to: 797

```

```

625 TTGGATGTGGCAGCATTTTCAGACGGCAGCTGCGCCAGCTGCTTA 576
      : : : : : : : : : : : : : : : : : : : : : : : :
228 ..AlaThrThrAlaAlaThrThrAlaAlaThrThrAlaAla... 242
      : : : : : : : : : : : : : : : : : : : : : : : :
575 CACACATGATTTTACGCTCGGTCAACGGCTCAATACACACATACG 526
      : : : : : : : : : : : : : : : : : : : : : : : :
243 ..... : : : : : : : : : : : : : : : : : : : : : :
525 TCTGAATACATCGACGCTTCTTGATCACAACACAGGCTCGCCGNA 476
      : : : : : : : : : : : : : : : : : : : : : : : :
248 rThrThrAlaAlaThrThrAlaAlaThrThrAlaAlaThrPro 265
      : : : : : : : : : : : : : : : : : : : : : : : :
475 GCGGATGTGTCATCGCATTTGAGAGAGATGGAGACGGCTCGCATMG 426
      : : : : : : : : : : : : : : : : : : : : : : : :
265 erGlySerThrSerThrGlyAlaSerThrSerThrProSerAlaSer 281
      : : : : : : : : : : : : : : : : : : : : : : : :
425 ACAGCA...GGGATTTTCTGACGAGGACGATACGACGAGTCCACAA 379
      : : : : : : : : : : : : : : : : : : : : : : : :
282 ThrAlaThrSerAlaThrProThrSerThrSerThrAlaAla.... 296
      : : : : : : : : : : : : : : : : : : : : : : : :
378 ACCGATGTGATCAGATTCGNNNNANNTCTMGCCCGCTTAAGTTGCA 329
      : : : : : : : : : : : : : : : : : : : : : : : :
297 ..... : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : :
328 ACCTTCGGGCGCGTAGCGTTGCACTGATTCGTCGCTGCTTCAACG 279
      : : : : : : : : : : : : : : : : : : : : : : : :
307 lAlaThrSerAla.....GlySerThrThrGlyAlaProThrSer 320
      : : : : : : : : : : : : : : : : : : : : : : : :
278 GCATTCAGACGACGACTGAGTACGCGCTTTCGCCGCGATGAGCGGC 229
      : : : : : : : : : : : : : : : : : : : : : : : :
321 ThrProThrThrAsp...ThrThrThrProSerGlyAlaThrAlaThr 336
      : : : : : : : : : : : : : : : : : : : : : : : :
228 GATTTGGCTGAACNGGCGGCTTAAC..... : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : :
336 rThrSerProGlySerThrThrValSerAlaSerThrThrSerAlaThr 353
      : : : : : : : : : : : : : : : : : : : : : : : :
200 .....ACACGCCGCGATCTTTTGTCTTCAACAGACACTGG 162
      : : : : : : : : : : : : : : : : : : : : : : : :
353 hrThrAlaPheThrThrGlySerThrThrSerProAspSerThrGly 369
      : : : : : : : : : : : : : : : : : : : : : : : :
161 CCTTTTTCAGCGCATCCGCTTCTGACTTTCATCMAGGGGCGCATACC 112
      : : : : : : : : : : : : : : : : : : : : : : : :
370 SerThrSerThrAlaGluProSerSerThrPheThrIleu.....ThrPr 384
      : : : : : : : : : : : : : : : : : : : : : : : :
111 GGCATATTCCTGCGCAGACGACGCTGCTGATGACGCGCCGCTCAT 62
      : : : : : : : : : : : : : : : : : : : : : : : :
384 oSerThrAlaThrProSer.....ThrAspAlaPheThrGlySerSer 399
      : : : : : : : : : : : : : : : : : : : : : : : :
61 AATGACT.....TCCTCCGCTGTCGCCGCGATGGGC 30
      : : : : : : : : : : : : : : : : : : : : : : : :
399 lAsrThrGluSerAspSerThrAspSerSerThrValProThrThrGly 415
      : : : : : : : : : : : : : : : : : : : : : : : :
seq_name: SwissProt_40:RNFC_HAEIN

```

```

seq_documentation_block:
ID RNFC_HAEIN STANDARD; PRT; 819 AA.

```

```

AC P71397;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Electron transport complex protein infc.
GN RNFC OR H11685.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
FX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,

```

```

RA Scott J.D., Shirley R., Liu L.-I., Glodok A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uitterback T.R., Hanna M.C., Nguyen D.T., Saudok D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RD Rd.";
RE Science 269:496-512(1995).
CC -1- FUNCTION: May be part of a membrane complex involved in electron
CC transport (By similarity).
CC -1- COPACTOR: Binds 2 4Fe-4S clusters (Potential).
CC -1- SUBUNIT: Composed of at least six subunits: rnfA, rnfB, rnfC,
CC rnfD, rnfE and rnfG (By similarity).
CC -1- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -1- SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC RNFC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32841; AAC23331.1; -.
DR HSSP: P00198; 2FBN.
DR TIGR: H11685; -.
DR InterPro: IPR001450; 4Fe4S_ferredoxin.
DR InterPro: IPR001949; Complex1_51K.
DR Pfam: PF01512; Complex1_51K; 1.
DR Pfam: PF00037; fer4; 2.
DR PROSITE: PS00198; 4FE4S_FERREDOXIN; 2.
DR KW Electron transport; Iron-sulfur; 4Fe-4S; Inner membrane;
KW Complete proteome.
FT METAL 378 378 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 381 381 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 384 384 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 388 388 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 417 417 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 420 420 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 423 423 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 427 427 IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ SEQUENCE 819 AA; 89283 MW; 541EAA2AEC6AE93 CRC64;

```

```

alignment_scores:
Quality: 116.50 Length: 468
Ratio: 0.552 Gaps: 17
Percent Similarity: 45.085 Percent Identity: 19.017

```

```

alignment_block:
US-09-303-518D-127 x RNFC_HAEIN

```

```

Align seq 1/1 to: RNFC_HAEIN from: 1 to: 819

```

```

70 CCGCTCATTCGAGAGTGGCGTGTGGCAGAAATATGCGGTATGCG 119
      : : : : : : : : : : : : : : : : : : : : : : : :
38 ProLeuGlyThrSprPheTyrlleProLeuGlyGlnHisLeuGlyThr 54
      : : : : : : : : : : : : : : : : : : : : : : : :
120 CCCCCTGATGAAAGTCAAGAGAGGCGATGCCGTCAAAAGAGCCAAAGTCG 169
      : : : : : : : : : : : : : : : : : : : : : : : :
54 rGlyAsnLeuLeuIleLeuGlyGlySprValLeuGlyGlyAlaAla 71
      : : : : : : : : : : : : : : : : : : : : : : : :
170 TGTTTGACAGCAAAAGATCCGGGCTGTGTTACCGCGCCGCTTTCA 219
      : : : : : : : : : : : : : : : : : : : : : : : :
71 eutThrGlyGlyAspGlyLeuArgMetLeuProValHisAlaProThrSer 87
      : : : : : : : : : : : : : : : : : : : : : : : :
220 GCGAAATGCGCGCGCATTCGCGGAGAAAGCGGCTACTTCAGTCGCT 269
      : : : : : : : : : : : : : : : : : : : : : : : :
88 GlyThrIle.....LysSerIleLeuProtyrIle 97

```


FT CARBOHYD 533 533 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 886 886 N-LINKED (GLCNAC...) (POTENTIAL)
 SQ SEQUENCE 1117 AA; 121063 MW; 87F13A07E42B0AD1 CRC64;

alignment_scores:

Quality: 110.00 Length: 309
 Ratio: 0.688 Gaps: 12
 Percent Similarity: 51.780 Percent Identity: 22.977

alignment_block:

us-09-303-518d-127/rev x YN96_YEAST

Align seg 1/1 to: YN96_YEAST from: 1 to: 1117

```

960 GTAGCGTCCCAAAATATCGTGGCGCTTGTGTATCGCGCGCTTCAATA 911
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
50 ValAlaGlnSerValValValValValValValValValValValVal 59
910 CCGAAGCGGAATACAGCGGCTTGTCTGCTGTC.....AACCAATGCG 879
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
59 lSerAlaSerHisSerProLeuSerValSerProLysAsnArgValS 76
869 CCGCGAGTAATTTGGCATACTTTCGACCCAAACGATGCGAGAGGCG 820
::: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
76 eAlaSerProLeuSerGlnAspSerAspSerValThrArgThra 92
819 TGGTTGTGACTTGAGAACACCCAAAGCAATCAGCGCTCGGTGTCA 770
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
93 .....ValGlnLeuSerLeuSerLeuSerValAlaSerGlnValS 107
769 GAGCGCGCTTGCACAAACGATCGGATGCGAATTCATCTTGATATG 720
: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
107 rGlnLysIleSerAlaGlnThrAsnAsn..... 116
719 ATGCTCAACACGCTTTGTTGTCACCGCGCTCATGAATGAATG 670
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
117 .....AspProValThrValSerAsnIle 124
669 CGTGCACACTCAACCGCGCGATGCGGCGCGCGCAATTCATGTTTGA 620
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
125 TTTAlaAsnAspAsnSerIleSerIleSerValIleAsnLeuSerS 141
619 TGTGTGACCATTTTCAGACGCGACGCTGCGCCAGCGCGCTTACACA 570
: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
141 rValSerGlyValAlaSerValMetProSerAlaSerThrMetArgLysV 156
569 TGGATTTTACGCTCGGTCAACGCGCTCAATACAGCANACNTGCTGANA 520
::: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
158 alThrThrLeuLeuSerGlnThrAlaSerThrSerThrSerThrLeuPh 174
519 ATGCTGCGCGCTTCTTGTATC.....ACAACCAAG 488
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
174 eSerSerSerLeuSerIleSerGlyThrGlnLeuAsnGlnIleThrLeuLeuT 191
487 GGTCTGCCCGCCNAGCGATGCTGTCATGCAATTCAGAGAGAGCGCAAC 438
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
191 hSerValSerIleSerGlyThrIleAspProLeuValThrGlnMetProSer 207
437 GGTCTGCATTCAGACGCGATTTTGTGAACGAGGAGNTACGACCGC 388
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
208 TyrSerSerGlnGlnThrIleIleProSerSerLeuThrSerAsnLys 224
387 AGTCCCAACACCGGATGATGATGATGATGATGATGATGATGATGATG 338
::: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
224 sThr.....IleThrIleSerValArgThrA 234
337 AGTTTCCCAACGCTTGGCGCGGTAGCGTTCGAACATCGATTTGCTGTG 288
::: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
234 snAlaIleThrAlaThrArgLys.....GluAspSerPheIleLeuIleSerThr 248
287 CCTTCAACGCAATTCACGACCGATGATGATGATGATGATGATGATG 238
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||:

```

```

249 ProAlaSer.....SerThrIleuPheTyProSerAs 259
237 GATGGCGCGGATTTTGCCTGAACNCGCGGTAACACGACCGCGCAT 188
::: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
259 nSerThrGlnAspLeuValGlnThrLeuAla...SerThrThrAlaSerP 275
187 NCTTTTGTCTTCAACACGACTTGGCTTTTGTGACGCGATCGCTTCC 138
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
275 roAlaTyProSerAsnArgThrGln.....IleThrLeuSerProSer 289
137 TTGACTTTTCATCNAGGCGCGCATACCGCATATCTTCCGCAACGACG 88
::: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
290 ValSerLeuTySerThrThrSerProIleTy.....ProSerAsnI 304
87 GACTCGGTATGACGCGCGCGTCA 63
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
304 eThrGluAsnGlySerSerProSer 312
seq_name: SwissProt_40: MURF_HAEIN
seq_documentation_block:
ID MURF_HAEIN STANDARD: PRF; 457 AA.
AC P45061.
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE UDP-N-acetylmuramoylalanine-D-glutamyl-2,6-diaminopimelate--D-alanyl-D-
DE alanyl ligase (EC 6.3.2.15) (UDP-MurNAc-pentapeptide synthetase)
DE (D-alanyl-D-alanine-adding enzyme).
GN MURF OR H1134.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-T., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uitterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd."
RL Science 269:496-512(1995).
CC -!- FUNCTION: INVOLVED IN CELL WALL FORMATION. CATALYSES THE FINAL
CC STEP IN THE SYNTHESIS OF UDP-N-ACETYLMURAMOYL-PENTAPEPTIDE, THE
CC PRECURSOR OF MUREIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanyl-D-
CC glutamyl-meso-2,6-diaminopimelate + D-alanyl-D-alanine = ADP
CC + phosphate + UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-6-
CC carboxy-L-lysyl-D-alanyl-D-alanine.
CC -!- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: BELONGS TO THE MURCDF FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U32793; AAC22789.1; -.
CC HSSP: P11880; IG64.
CC
CC TIGR: H11134; -.

```


RP SEQUENCE FROM N.A. AND SEQUENCE OF 1-6.
RC STRAIN-ATCC 33303 / B10;
RX MEDLINE-98151232; PubMed-9492268;
RA Jouanneau Y., Jeong H.-S., Hugo N., Meyer C., Willison J.C.;
RT "Overexpression in Escherichia coli of the rnf genes from Rhodobacter
RT capsulatus -- characterization of two membrane-bound iron-sulfur
RT proteins.";
RL Eur. J. Biochem. 251:54-64(1998).
[3]
RP SUBCELLULAR LOCATION.
RC STRAIN-SB1003 / St Louis;
RX MEDLINE-9729785; PubMed-9154934;
RA Kumagai H., Fujiwara T., Matsubara H., Saeki K.;
RT "Membrane localization, topology, and mutual stabilization of the
RT rnfABC gene products in Rhodobacter capsulatus and implications for a
RT new family of energy-coupling NADH oxidoreductases.";
RL Biochemistry 36:5509-5521(1997).
CC -1- FUNCTION: Required for nitrogen fixation. May be part of a
CC membrane complex functioning as an intermediate in the electron
CC transport to nitrogenase. Stabilizes rnfB.
CC -1- COPOLYMER: Binds 2 4Fe-4S clusters (Potential).
CC -1- SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
CC rnfD, rnfE and rnfG.
CC -1- SUBCELLULAR LOCATION: Inner membrane-associated (Probable).
CC -1- INDUCTION: Expression is reduced under iron-limiting conditions.
CC -1- SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC RNFBC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X72888; CA51399.1; -
DR EMBL: Y11913; CA72670.1; -
DR InterPro: IPR001450; 4Fe4S_ferredoxin.
DR InterPro: IPR001949; Complex_51K.
DR Pfam: PF05151; Complex_51K; 1.
DR PROSITE: PS00198; 4Fe4S_FERREDOXIN; 2.
KW Nitrogen fixation; Electron transport; Iron-sulfur; 4Fe-4S;
KW Inner membrane.
FT METAL 381 381 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 384 384 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 387 387 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 391 391 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 420 420 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 423 423 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 426 426 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 430 430 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT DOMAIN 498 503 POLY-A
SQ SEQUENCE 519 AA; 55587 MW; E5451EB2A3FA6026 CRC64;

alignment_scores:
Quality: 109.50 Length: 478
Ratio: 0.547 Gaps: 19
Percent Similarity: 41.841 Percent Identity: 19.456

alignment_block:
US-09-303-518D-127 x RNFBC_RHOCA ..
Align seg 1/1 to: RNFBC_RHOCA from: 1 to: 519

```

70 CCGGTCATTACCGAAGTCCGCTGCTTGGCGAAGATATGCGCGATGCGG 119
   |||::: :::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
40 PrometProAlaLeuIleArgLeuProLeuGlnGlnHisIleGlyAlaI 56
120 CCGGTCATTACCGAAGTCCGCTGCTTGGCGAAGATATGCGCGATGCGG 169
   |||::: :::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
56 uAlaGluProIleValIleValIleValIleValIleValIleValIle 73

```

```

170 TGTGTGAAGCAAAAGNATCCGGCGGTGTTATCCGGCCGCGTTC 219
   :::: :::: ::::: ::::: ||| ||| ||| ||| ||| ||| |||
73 IeAlaIleValIleValIleValIleValIleValIleValIleVal 89
220 GGGCAAAATCCGCGCATC.....CATCCGGCGCGA 251
   |||::: |||::: |||::: |||::: |||::: |||::: |||:::
90 GlyArgValIleValIleValIleValIleValIleValIleValI 106
252 CCGGTCATTACCGAAGTCCGCTGCTTGGCGAAGATATGCGCGATG 297
   ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
106 uProVal...ProIleIleThrIleArgProArgIleValIleVal 122
298 .....ATCGATTCGAACGCTACCGCGCGCGAAGCGTTG 330
   ::::: |||::: |||::: |||::: |||::: |||::: |||:::
122 IyProHisLeuProAlaGluLeuArgProGlnAlaIleProGlnIle 138
331 GCA..... 333
   |||
139 AlaAlaGlnValAlaAlaIleValIleValIleValIleValIle 155
334 .....AAGTTAAGCGCGCGAAGATATGCGCGATC 364
155 eProSerAlaValIleValIleValIleValIleValIleValIle 172
365 TGAATCCAAATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 408
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
172 eulleIleAsnGly.....AlaGluGluGluProIleValIleThr 184
409 .....AGCAAAATCCCTGCGCGATGCGCGACCGTTGCC..... 444
   ::::: |||::: |||::: |||::: |||::: |||::: |||:::
185 CysAspArgPargLeuMetArgGluArgAlaGluIleAlaIleAla 201
445 .....ATCTGCTCATGCGA 460
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
201 eGlyIleMetAlaIleArgAlaLeuGluValIleValIleValIleVal 218
461 TGGACCAAAATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 510
   ::::: |||::: |||::: |||::: |||::: |||::: |||:::
218 IuseArgIleValIleValIleValIleValIleValIleValIle 227
511 GNCAGNATTTGACAGCAGANTGCTGCTGCTGCTGCTGCTGCTGCG 560
   ::::: |||::: |||::: |||::: |||::: |||::: |||:::
228 MetIleArgIleValIleValIleValIleValIleValIleValI 239
561 TAAATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 610
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
240 .LysIleHisVal.....ValProIleGlnIleVal 249
611 CTGGCAACATGGAACACATGATTC.....GGCGCGCGCAT 648
   ::::: |||::: |||::: |||::: |||::: |||::: |||:::
249 rometGlySerGlnIleValIleValIleValIleValIleValI 265
649 CCGGCGGCTTGAAGTGGACGACATTCATTCATTCATTCATTCATTC 698
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
266 ProAla.....ArgAlaLeuThrAl 272
699 AAGCAAAATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 748
   ||| ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
272 aAspLeuGlyValIleValIleValIleValIleValIleValIle 289
749 GTTGTGTTGCAACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 798
   |||::: |||::: |||::: |||::: |||::: |||::: |||:::
289 euaIleValIleValIleValIleValIleValIleValIleValI 305
799 GGTTCGACGACCAACACGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 848
   ||| ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
306 GlyHisIleGlyIleArgPargProAlaAsnIleValIleValIle 322
849 AGTATCGCAATTAAGTGC.....GGCGAATTCGTTGACGACGAC 892
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
322 oValSerIleIleIleAlaHisGlyIleValIleValIleValIle 339

```

```

893 GCGTGAATTCGGTTCGGTATTTGAACGGCCGATTAACAGAGCCGAC 942
      |||::: |||::: |||
339 rgleuleuleuglyglyprometmetglymetproleulin..... 352
      |||::: |||::: |||
943 GATTATTTGGAGCGCTACCAATCAGATTTCCGTTATTCAGAGAGCGC 992
      |||::: |||::: |||
353 .....AsnProArgValProValValLysGly..... 361
      |||::: |||::: |||
993 CAGCAAGAGCTGTTGGCTGGGTTGGCCGCGACCGACAAATACCTCA 1042
      |||::: |||::: |||
361 ..... 361
1043 TCACGCGTAGACCCCTGCGCATTTCCGAAACAACTCTCAAGTTC 1092
      |||::: |||::: |||
362 .....ThraSngLYleLeuAlaLeu 368
      |||::: |||::: |||
1093 ACAGACGCCGTCACAGGTGGCGACCGCCGATG.....GTGCGGATTGG 1136
      |||::: |||::: |||
369 ThrAlaAlaGluThrProGluAlaLysThrMetProCysIleArgCysG 385
      |||::: |||::: |||
1137 TACTTAGAGCGCGTAATGCCGCTAGACATCTGCTACCTGCTTTTGC 1186
      |||::: |||::: |||
385 YArgCysValGlnGlyCysProValGlyLeuThrProhGluLeuAsnA 402
      |||::: |||::: |||
1187 GCGATTTAATGTCGCGGATACGACAGCCGCGAAGCATTTGGCT..... 1230
      |||::: |||::: |||
402 IaArgLYleHisAlaGlyAspLeuGluGlyAlaLysValGlyLeuMet 418
      |||::: |||::: |||
1231 ...TGCTGGAAATTGGACGAAGAAGACCTGCTTGGACGCTTCGCTG 1277
      |||::: |||::: |||
419 AspySLeuAla.....CysGlyCysCysSeryIraSncY 430
      |||::: |||::: |||
1278 CCGCGCAATATACGAATANGCCGCTGTTGCGT 1311
      |||::: |||::: |||
430 sProAlaAsnLeu.....ProLeuValGln 438
      |||::: |||::: |||
seq_name: SwissProt_40:YGY3_HALSQ
seq_documentation_block:
ID YGY3_HALSQ STANDARD; PRT; 437 AA.
AC P21561.
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 50.6 kDa protein in the 5' region of GYRA and GYRB (ORF
DE 3).
OS Haloferax sp. (strain Aa 2.2).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Haloferax.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91100352; PubMed=1846146;
RA Holmes M.L., Dyall-Smith M.L.;
RT "Mutations in DNA gyrase result in novobiocin resistance in
RT halophilic archaeobacteria.";
RL J. Bacteriol. 173:642-648(1991).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL_M38373; -; NOT_ANNOTATED_CDS.
DR PIR_C39135; C39135.
KW Hypothetical protein.
SQ SEQUENCE 437 AA; 50626 MW; B5B99AZAF3892BEF CRC64;

```

alignment_scores: Quality: 108.00 Length: 480

Ratio: 0.554 Gaps: 27
Percent Similarity: 40.625 Percent Identity: 23.958

alignment_block:

us-09-303-518d-127 x YGY3_HALSQ ..

Align seg 1/1 to: YGY3_HALSQ from: 1 to: 437

```

12 CAAAAAGGCTTAACCTGCCATCGCGGCGACGCGCAAGCATTT 61
      |||::: |||::: |||
48 GlnAspAlaArgLysProAlaArgValProGlnLeu..... 62
      |||::: |||::: |||
62 ATACGCGCGCGCTATACGAGATGCGGTGGTGGCGAAGATATGCG 111
      |||::: |||::: |||
63 ...ArgGlyArgAspPheAlaLeuArgArgAlaAspArgVal... 77
      |||::: |||::: |||
112 GGTATGCGCGCCCTNGATGAAGTCAAGAGAGCGATGCGTCAAAAAG 161
      |||::: |||::: |||
77 LnhIsvAlProLeu.....ArgGlyArgHisProAlaArgValArg 90
      |||::: |||::: |||
162 ...CCAAGTCT.....GTTGAAG 178
      |||::: |||::: |||
91 ValProGlnArgAspGlnAspGlyAlaProArgArgHisLeuThr 107
      |||::: |||::: |||
179 ACAAAAGNATCCGGG...CGTGGTGTAAACCG..... 209
      |||::: |||::: |||
107 GATGATGValGlyGlyHisArgGlyArgAsnArgHisAlaGlyAspArg 124
      |||::: |||::: |||
210 .....GCCNGTTGAGGCAAAATGCGCGCATTCATCG 242
      |||::: |||::: |||
124 rGAlaProGlyAlaAspSerArgLeuArgGlnHisGlnHisProArg 140
      |||::: |||::: |||
243 CCGCGAAAGCGCGTACTGACGCGTGTGATGGCGGTGAAGCAAGC 292
      |||::: |||::: |||
141 GlyArgHisAlaSerAspArgValGlnAspGlyAla.....HisPro 155
      |||::: |||::: |||
293 ACGAATCGAGTTGCAACGCTACGCGCGCGAAGCGTTGGCAACTTAAC 342
      |||::: |||::: |||
155 GATGlnArgLeuArgGlu...GlnProArgHisAlaGlyArgProArg 171
      |||::: |||::: |||
343 GCGGANGAANTNNGNCAATGATCAATCGGTTGTGATGCGCT 392
      |||::: |||::: |||
171 rGArg..... 172
      |||::: |||::: |||
393 GCGTACCGCTCGTTGACAAATCCCTCGCTGATGCCGAGCC..... 437
      |||::: |||::: |||
173 .....GlnProProArgArgGlyArgSerArgG 182
      |||::: |||::: |||
438 .....GTTGCCATCTGTCATATCGATGACACCAATCCGCTNGCG 480
      |||::: |||::: |||
182 yThrHisArgArgHisLeuArgGlnAlaPro..... 192
      |||::: |||::: |||
481 GCAGACCGCTGT.....GGTTGATCAAAAGACCGGANGATTTACG 524
      |||::: |||::: |||
193 ...ArgProAlaValArgGlyProAspGluAspGlnAlaArgGlnubheArg 208
      |||::: |||::: |||
525 ACAGANTNCTGCTGATTTGACCGCTTGGACGAGCGTAAATCCATGTGT 574
      |||::: |||::: |||
208 ..... 208
      |||::: |||::: |||
575 GTAAGCAGCTGGCGCAGACGCGCGTGTGAATATGCTCCAAATCGAA 624
      |||::: |||::: |||
209 ...GlyProArgHisArgArgGlu..... 215
      |||::: |||::: |||
625 ACACATGAATTGCGCGCGCGCATCGCGCGG.....TTT 659
      |||::: |||::: |||
216 .....ArgHisProProThrAlaArgAspValLeuArgGlyGlu 228
      |||::: |||::: |||
660 GAGTGCACGACATTCATTCATGAGCCGCTGCGTGAACAAACACCG 709
      |||::: |||::: |||
229 ProGlyHisGlyAspGlyHisLeuGluGlyArgArgGlyArgProArg 245
      |||::: |||::: |||
710 TTTGACCATCAATTATCAAGATGTAATTGCCATGAGCGTTGTTGCA 759

```



```

491 ACAGGCTCGCCGACGATGTCATCGCATTCAGCAAGATGCG 442
      ::::::::::: ||| :::::::::::
68 SerSerSerLeuThrSerSerThrSerAlaSerLeuValSerHis 84
      ::::::::::: ||| :::::::::::
441 GACAGGCTCGCCGACGATGTCATCGCATTCAGCAAGATGCG 392
      ::::::::::: ||| :::::::::::
84 rSerSerSerThrSerLeuAlaSerLeuSerPheThrSer 97
      ::::::::::: ||| :::::::::::
391 GCGCAGTCCACAAACCGATTCAGATTCGCAATTCGTCGCGC 342
      ::::::::::: ||| :::::::::::
98 ..... PheSerPheSerSerSerSerThr 106
      ::::::::::: ||| :::::::::::
341 CTTAAGTTGGCCACGCTTCGCGCGCGCTTCGATTCGATTCGTC 292
      ::::::::::: ||| :::::::::::
107 SerSerSerSerAlaSerSerSerSerSerSerSerPhe 123
      ::::::::::: ||| :::::::::::
291 GTTGGCTTCACGCGCATTCAGCGGCTTCGATTCGCGCGC 242
      ::::::::::: ||| :::::::::::
123 rLeSerSerThrSerAlaThrSerGluSerSerThrSerThrGln 140
      ::::::::::: ||| :::::::::::
241 GATGATGCGCGCGCTTCGCTGAACGCGCGGTAAACACACGCGC 192
      ::::::::::: ||| :::::::::::
140 hrSerThrSerSer..... SerSerSerLeuSerSerThrPro 152
      ::::::::::: ||| :::::::::::
191 GGATTCCTTTTGTCTCAACAGCAGCTTCGCTTCGATTCGCGC 142
      ::::::::::: ||| :::::::::::
153 ..... SerSerSerSerSerSerSerThrThrThrThrAlaThr 165
      ::::::::::: ||| :::::::::::
141 TTCCTGCTTCATTCAGCGCGCATTCGCGCATTCGCGCAACA 92
      ::::::::::: ||| :::::::::::
165 SerSerSerSerSerSerSerSerThrThrThrAlaThrAlaThr 181
      ::::::::::: ||| :::::::::::
91 ACAGGCTCGCCGATTCAGCGCGC 69
      ::::::::::: ||| :::::::::::
181 hrLeThrSerLeuThrThrThrThrThrThrThrThrThrThr 188
      ::::::::::: ||| :::::::::::
seq_name: SwissProt_40:AGAL_YEAST

seq_documentation_block:
ID AGAL_YEAST STANDARD; PRT; 725 AA.
AC P32323;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE A-agglutinin attachment subunit precursor.
OS AGAL OR YNR044W OR N3431.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Pohl T.M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MEDIATES THE CELL SURFACE ATTACHMENT OF THE A-AGGLUTININ
CC SUBUNIT. S-CEREVISIAE A AND ALPHA CELLS EXPRESS THE COMPLEMENTARY
CC CELL SURFACE GLYCOPROTEINS A-AGGLUTININ AND ALPHA-AGGLUTININ,
CC RESPECTIVELY, WHICH INTERACT WITH ONE ANOTHER TO PROMOTE CELLULAR
CC AGGREGATION DURING MATING.
CC -1- SUBUNIT: CONTAINS AT LEAST A BINDING SUBUNIT DISULFIDE-LINKED TO
CC A CORE SUBUNIT.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC AND/OR BY CARBOHYDRATE-MEDIATED COVALENT CROSS-LINKS (POSSIBLE).
CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M60590; AAA34382.1; -
DR EMBL: 271659; CAN96325.1; -
DR PIR: S17031; S17031.
DR PIR: A41258; A41258.
DR SGD: S0005327; AGAL.
KW Glycoprotein; Cell adhesion; Signal; GPI-anchor; Repeat;
KW Pheromone response.
FT SIGNAL 1 22
FT CHAIN 23 725
FT DOMAIN 53 493
FT REPEAT 53 149
FT REPEAT 395 493
FT REPEAT 182 307
FT FT 182 307
FT REPEAT 182 188
FT REPEAT 189 195
FT REPEAT 196 202
FT REPEAT 203 209
FT REPEAT 210 216
FT REPEAT 217 223
FT REPEAT 224 230
FT REPEAT 231 237
FT REPEAT 238 244
FT REPEAT 245 251
FT REPEAT 252 258
FT REPEAT 259 265
FT REPEAT 266 272
FT REPEAT 273 279
FT REPEAT 280 286
FT REPEAT 287 293
FT REPEAT 294 300
FT REPEAT 301 307
FT REPEAT 307 307
SQ SEQUENCE 725 AA; 73353 MW; 70420C853B0B01P8 CRC64;

alignment_scores:
Quality: 107.50 Length: 254
Ratio: 0.846 Gaps: 11
Percent Similarity: 50.000 Percent Identity: 24.803

alignment_block:
US-09-303-518D-127/rev x AGAL_YEAST ...

Align seg 1/1 to: AGAL_YEAST from: 1 to: 725
618 GTTGGCAGCATTTTCAGACGCGACGCTGCGCGCATTCACACAT 569
      ||| ||| ||| ::::::::::: ||| |||
124 ValThrSerLeuSerThrThrThrThrThrThrThrThrThr 139
      ::::::::::: ||| :::::::::::
568 GGAATTTTACGCTCGCTCAACAGCGCTCAATAC.....AGANA 531
      ::::::::::: ||| :::::::::::
139 hrAlaLeuSerSerLeuSerGluValGlyThrThrThrValAlaSer 155
      ::::::::::: ||| :::::::::::
530 CNTGCTGAATTCGTCGCGCGCGCTTCGATTCGATTCGATTC 498
      ::::::::::: ||| :::::::::::
156 SerAlaIleGluProSerSerAlaSerIleIleSerProValThr 172
      ::::::::::: ||| :::::::::::
497 .....ACAA 494
      ::::::::::: ||| :::::::::::
172 rLeuSerSerThrThrThrSerSerSerProThrThrThrSerLeuSer 189
      ::::::::::: ||| :::::::::::
493 CCACAGGCTCGCCGACGATTCGTCATTCGATTCGACGACGATG 444
      ::::::::::: ||| :::::::::::
189 hrSerThrSerProSerSerThrThrThrThrThrThrThrThr 205
      ::::::::::: ||| :::::::::::

```


Align seg 1/1 to: M121_HUMAN from: 1 to: 1229

```

1115 TCGCCACCGCTTGAGCGCTGCTGTAAC.....TTGAGAG 1081
      |||||
561 SerProthrinSerLeuAlaProSerThrAsnProLeuLeuGlu 577
1080 TTTGTTTTCAGGAATGCGGAGGTCGACGCGTGAGTAAATTGT 1031
      |||||
577 rLeuLysLysMetGlnThrProSerLeuProGlySerA 594
1030 CCGGCTGCGGCGAACCCGACGACG.....TCTTG 996
      |||||
594 laGlyAlaAlaThrThrGlnAlaLeuSerProLysThrProSerLeu 610
995 CTGGGCGCTTCTTGATACGGAATCTGATGTGTAGCGTCCCAATA 946
      |||||
611 LeuProProLeuGlyLeuSerGln..... 618
945 ATCGTGGCGCCCTTGTTGTAATCCGCG.....TTCATACGGAAC 905
      |||||
619 ..SerGlyProProGlyLeuLeuProSerProSerPheAspSerLysP 634
904 CGGAATACAGCGGTTGCT.....CGTCACCAATTGCGCCGCA 864
      |||||
634 roProthrinLeuLeuGlyLeuLeuProAlaProSerMetAlaProAla 650
863 GTATTTCGATCTTCGACCCCAAAACGTCGCAAGAGCGGTGTTT 814
      |||||
651 Thr.....AspThrLysAlaProProThrLeuGlnAlaGlnThrAla 665
813 GTTGACTTGAGAACCA.....CCCAAGCAATCGCGCT 779
      |||||
665 r.....LysProGlnAlaThrSerAlaProSerProAlaProLysG 679
778 CGGTTTCAGACGCGCTTGCAACAACATCCGATG..... 741
      |||||
679 InSerPheLeuPheGlyThrGlnAsnThrSerProSerProAlaAla 695
740 ...GCAATTACATCTTGATTAATGATGTCACAAACGTTTGTTCAC 694
      |||||
696 ProAlaAlaSerSerAlaSerProMetPheLysProLeuThrAlaPr 712
693 GACCGGCTCATGAATGAATGTCGTCGACCTCAACCG..... 654
      |||||
712 opProLysSerGlnLysGlnGlyProThrProGlyProSerValThrA 729
654 ..... 654
729 laThrAlaProSerSerSerLeuProThrThrSerThrAla 745
653 .....GCCGATGGGCGCGCCGCAATTCATGTGT 625
      |||||
746 ProThrPheGlnProValPheSerSerMetGlyProProAlaSer...Va 761
624 TTGATGTTGGCAGCATTTTCACAGCGACGTCGCGCAAGCTGCTTAC 575
      |||||
761 lProLeuProAlaProPhePheLysGlnThrThrProAlaThrAla 778
574 ACACATGATTTTACGTCGTCACAAACGCGCTCAATACGACNAACNTGT 525
      |||||
778 rothr..... 779
524 CTGAATTCNTGNCGGCTTCTTGATACACACAGGTCGCCGAC 475
      |||||
780 .....ThrThrAlaProLeuPheThrGlyLeuAlaSerAlaThrse 793
474 CGGATGTGTCCATGTCATGACGAAGATGGGAACGCTGCGCATC 426
      |||||
793 rAlaValAlaProLeuThrSerAlaSerProSerThrAspSerAlaSerL 810
425 .....ACGGCAGGATTTGCTGACGACGAGGTCGACGACGACGAC 381
      |||||
810 ysrProAlaPheGlyPheGlyLysAsnSerValSerSerSerVal... 825

```

```

380 AACCGGATTGATCAGATTGCNNCNCNNTTCGCGCTTAAGTTGC 331
      ::
826 SerThrThrThrSerThrAlaThrAlaAlaSerGlnProPheLeuPheG1 842
330 C.....AACGCTCGGGCGCGCTGACGCTTGCACACTCGATTTCGTTG 288
      ::
842 yAlaProGlnAlaSerAlaAlaSerPheThrProAlaMetGlySerL1eP 859
287 .....CCTTCACGCAATCAGCAGCACTGAAGTACGCC 252
859 heGlnPheGlyLysProProAlaLeuProThrThrThrValThrThr 875
251 TTTTCGCCCGCATGATGGCGGCGATTTTGCCTGAACNGCGCGGTAA 202
876 PheSerGlnSerLeuProThrAlaValProThrAlaThrSerSerAl 892
201 CACCAAGCCCGGATNCTTTTGTCTCAACAGCACT...TGGCTTTT 155
892 aAlaAspPheSerGlyPheGlySerThrLeuAlaThrSerAlaProAla 909
154 TGACGCGATCGCTTCTTGATTCATCNCAGGCGCATACCGGATAT 105
909 hrSerSerGlnProThrLeuThrPheSerAsnThrSerThrProThrPhe 925
104 TCTTCGCCACAGCAGCGCACTCGCTAATGACGCGCGCTCATTAATGAC 55
926 Asn11eProPheGlySerSerAlaLysSerProLeuProSerTyProG1 942
54 TTTCTCGGCTGTCGCCGCGATGGC 30
942 yAlaAsnProGlnProAlaPheGly 950

```

seq_name: SwissProt_40:ATPB_PYLII

seq_documentation_block:

```

ID ATPB_PYLII STANDARD; PRT; 481 AA.
AC P26532;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ATP synthase beta chain (EC 3.6.3.14).
GN ATPB.
OS Pyralia litoralis.
OG Chloroplast.
OC Eukaryota; stramenopiles; Phaeophyceae; Ectocarpales;
OC Acinetosporaceae; Pyralia.
OX NCBI_TaxID=2885;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92216062; PubMed=1532750;
RA Jouanin S., Kerbourn H.C., Kloeareg B., Loiseaux-De Goer S.;
RT "Nucleotide sequences of the atpB and the atpE genes of the brown
RL alga Pyralia litoralis (L.) Kjellm."
CC Plant Mol. Biol. 18:819-822(1992).
CC -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC
CC SUBUNIT.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +
CC H(+)(Out).
CC -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -!- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```



```

FT CARBOHYD 104 104 O-LINKED (GALNAC. . .) (MUCIN TYPE)
FT CARBOHYD 106 106 O-LINKED (GALNAC. . .) (MUCIN TYPE)
FT CARBOHYD 107 107 O-LINKED (GALNAC. . .) (MUCIN TYPE)
FT CARBOHYD 108 108 O-LINKED (GALNAC. . .) (MUCIN TYPE)
FT CARBOHYD 110 110 O-LINKED (GALNAC. . .) (MUCIN TYPE)
FT CARBOHYD 114 114 O-LINKED (GALNAC. . .) (MUCIN TYPE)
FT CARBOHYD 117 117 O-LINKED (GALNAC. . .) (MUCIN TYPE)
FT CARBOHYD 123 123 O-LINKED (GALNAC. . .) (MUCIN TYPE)
FT CARBOHYD 124 124 O-LINKED (GALNAC. . .) (MUCIN TYPE)
FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 547 547 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 917 917 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 985 985 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1002 1002 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1068 1068 N-LINKED (GLCNAC. . .) (POTENTIAL)
SO SEQUENCE 1150 AA: 109615 MW: 368685D29D7F5A CMC64;

```

alignment_scores:

```

Quality: 105.00 Length: 450
Ratio: 0.507 Gaps: 22
Percent Similarity: 46.000 Percent Identity: 24.667

```

alignment_block:

US-09-303-518d-127/rev x APMU_PIG

Align seg 1/1 to: APMU_PIG from: 1 to: 1150

```

1325 GTTCCACACCTTACGACAGCGGGCCNATTGTTTGGCCGGCA 1276
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
242 VAlThrGluThrAlaArgProSerValAlaGlySerGlyThrGly.. 257
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1275 GACGAGCTGCACAAAGAGGAGTCTTCGTCAATTCGAACCAACCA 1226
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
258 .ThrValSerGlyAlaSerGlySerThrylSerSerSerGlySerProG 274
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1225 ATGCTGGCGCGCTGTGCTATCCGCGACGATTAAATCGCGAAAACAGG 1176
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
274 lYAlaThrGlyAlaSerIleGlyGlnProGluThrSerArgIleSerVal 290
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1175 GAGGCG.....AGATGTCAGCGGCACTTACGCGCTGCTA 1141
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
291 AlaglySerSerGlyAlaProAlaValSerSerGlyAlaSerGlnAla.. 306
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1140 AGTACCAATCGCACCATGCGCGGCGCCACCGTTCAGCGCTGCTGTA 1091
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
307 .....AlaglyThrSerGlyAlaGlyProGlyThrThrAla..... 318
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1090 ACTTGAGAGATTGTTTTCAGGAATGCGGAGGCTCGTACGCGGTGATG 1041
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
319 .....SerSerValGlyValThr 324
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1040 GAG.....TATTGTCGGCGTGGCGCCCAACCCGCGCAACAG 1003
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
325 GluThrAlaAlaGProSerValAlaGlySerGlyThrThrGlyThrValSe 341
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1002 CTCTTGGTCGCGGCTTCTTCGATACGGAATCTGATTGTTGTTAGCTC 953
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
341 rGlyAlaSerGlySerThrGlySerSer..... 350
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
952 CCAATTAATATGTCGCGCGCTGTGTAATGCGCGCTTCATATACGAACCG 903
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
351 .....SerGlySerProGlyAlaThrGlyAlaSerIleGlyGlnPro 364
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
902 GAAATACAGCGGTGTCT.....GCGTCACCAATTCGCGCGCGAGTAAT 859
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
365 GluThrSerArgIleSerValAlaGlySerSerGlyAlaProAlaValSe 381
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
858 TTGGCACTTTCGCACCCCAACAGGTACGACGAGCGCTGTTGTTGA 809
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
381 rSerGlyAla..SerGlnAlaAla..GlyThrSerGlyAla..... 393
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
808 CTTGAGAACCAACCCCAACCAATACGCGG.....CTCGGTGTTTC 771
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
394 .....ThrThrSerIleGlyAlaGlyThrSerGlyValGlyPhe 408
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
770 AGACGCGCTGTGC.....AACCAACGTCGATGCGCATTCACCTTG 727
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
408 ysrThrGluAlaThrThrPheProGlyGluGlnGluThrThrArgValGly 424
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
726 ATATATGATGTCCAAACGTTTGTGTCACCGACCGCTC.....AA 683
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
425 lLeAlaThrGlyThrThrGlyIleValSerArgIleThrLeuGluProG 441
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
682 TGAATGAATGTGCGTGCACCTCAACCC...GGCCGATGCGGCGCGCG 636
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
441 yserTyrAsnThrGluAlaThrThrSerIleGlyArg..... 453
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
635 AATTCATGTGTTCGATGTTGCGACGATTTTCAGACGCGCAGCTGCGCC 586
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
453 ..... 453
585 AGCTGCTTACACACATGATTTTACGCTCGTCACACG.....GC 545
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
454 SerGlyThrThrThrThrAspLeuProGlyGlyThrThrIleValLeuPr 470
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
544 TCAATACGACANACNTGCTGTAATCNCGCGCTTCTTGATCACA 495
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
470 oglyPheSerHisSerSerGln..SerSerIlyProGlySerSerValThr 486
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
494 ACCACAGGCTTCGCGCGACGATGTTGTCATGCA..... 456
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
487 ThrProGlySerProGlySerGlySerGlyThrGlyThrSerGlyLubh 503
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
455 .TTGACGAGATGCGGACGCTCGGCAATCGACGCGAGGATTTTCTCA 407
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
503 eserThrThrValIleSerIlySerSerHisThrGluAlaThrThrPhe 520
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
406 ACGACGAGTACGACGCGAGTCACACAAACCGGATGATGATTCNN 357
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
520 lGlyGlySerGlySerPro.....Gly 527
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
356 CNAANTTCNTCGCCCTTAAGTTGCCAACGCTTCGCGCGCGATGCTTC 307
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
528 ThrGlySerArgProGlyThrThrGlyGluLeuSerGly.....Th 541
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
306 GAACCTGATTCGTCGTTCGCTTCACAGCGCAATACGACGAGTGAAGA 257
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
541 rThrIleAlaSerGlyAsnAlaThrThrGluAlaThrThrSerThrGlu 558
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
256 CGCGCTTTTCGCGCGATGATGCGCGATTTTGGCTGAACNGCGCG 207
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
558 hrArg.....IleGlyProGlnThrGlyAla 566
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206 GTAACACACGCGCGGATTCCTTGTCTTCAACAGCACTTGCCCTTT 157
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
567 GluThrThrValProGlySerGlnValSerGlySerGlyThrGlyThrSe 583
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
156 TTGACGCGCATGCGCTTCCTTACTTTCATCNAAGGGCGCATACGGCAT 107
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
583 rGluAlaValSerAsn.....ProAla 591
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
106 ATTCTTCGCCAACGACGACTTCGTAATGACGCGCGCTCA 63
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
591 leAlaSerGlySerSerSerThrGlyThrThrSerGlyAlaSer 605
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: SwissProt_40: FAS1_SCHPO
seq_documentation_block:
ID FAS1_SCHPO STANDARD; PRT; 2073 AA.
AC Q9UG0; P78799;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fatty acid synthase subunit beta (EC 2.3.1.86) [includes: 3-
hydroxypalmitoyl-[acyl-carrier-protein] dehydratase (EC 4.2.1.61);

```

DE Enoyl-[acyl-carrier-protein] reductase [NADH] (EC 1.3.1.9); [acyl-carrier-protein] acetyltransferase (EC 2.3.1.38); [acyl-carrier-protein] malonyltransferase (EC 2.3.1.39); S-acyl fatty acid synthase thioesterase (EC 3.1.2.14)].

GN FAS1 OR SPAC926.09C.

OS Schizosaccharomyces pombe (fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomyces.

OX NCBI_Taxid:4896;

RN [1]

RP SEQUENCE FROM N.A. AND SUBUNIT.

RX MEDLINE=98359895; Pubmed=9693066;

RA Niwa H., Katayama E., Yanagida M., Morikawa K.;

RT "Cloning of the fatty acid synthetase beta subunit from fission yeast, coexpression with the alpha subunit, and purification of the intact multifunctional enzyme complex."

RT multifunctional enzyme complex."

RL protein Expr. Purif. 13:403-413(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RA Wood V., Rajandream M.A., Barrell B.G., Ramsperger U., Pohl T.;

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 1725-2073 FROM N.A.

RX STRAIN=PR745;

RX MEDLINE=98162722; Pubmed=9501991;

RA Yoshida S., Kato K., Nakai K., Okayama H., Nojima H.;

RT "Identification of open reading frames in Schizosaccharomyces pombe cdnas."

RT DNA Res. 4:363-369(1997).

RL -1 FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH. THE BETA SUBUNIT CONTAINS DOMAINS FOR: [ACYL-CARRIER PROTEIN] ACETYLTRANSFERASE AND MALONYLTRANSFERASE, S-ACYL FATTY ACID SYNTHASE THIOESTERASE, ENOYL-[ACYL-CARRIER PROTEIN] REDUCTASE, AND 3-HYDROXYPALMITOYL-[ACYL-CARRIER PROTEIN] DEHYDRATASE.

CC -1 CATALYTIC ACTIVITY: ACETYL-COA + N MALONYL-COA + 2N NADPH =

CC LONG-CHAIN FATTY ACID + (N+1) COA + N CO(2) + 2N NADP(+).

CC -1 CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA + acetyl-[acyl-carrier protein].

CC -1 CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA + malonyl-[acyl-carrier protein].

CC -1 CATALYTIC ACTIVITY: (3R)-3-Hydroxypalmitoyl-[acyl-carrier protein] = 2-Hexadecenoyl-[acyl-carrier protein] + H(2)O.

CC -1 CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NAD(+) = trans-2,3-dehydroacyl-[acyl-carrier protein] + NADH.

CC -1 CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-carrier protein] + oleate.

CC -1 SUBUNIT: FATTY ACID SYNTHETASE ARE [ALPHA(6)BETA(6)] HEXAMERS OF TWO MULTIFUNCTIONAL SUBUNITS (ALPHA & BETA).

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AB010274; BAA36384.1; -

DR EMBL; AL110469; CAB54157.1; -

DR EMBL; D89148; BAA13810.1; -

DR InterPro: IPR001227; Acyltransf.domain.

DR InterPro: IPR003965; Fatty acid synth.

DR InterPro: IPR002539; MacC.dehydratase.

DR Pfam; PF00698; Acyl.transf.1.

DR Pfam; PF01575; MacC.dehydratase.1.

DR PRINTS; PR01483; FASYNTHASE.

KW Fatty acid biosynthesis; Multifunctional enzyme; Oxidoreductase;

KW Transferase; Hydrolase; Lyase; NAD; NADP.

FT DOMAIN 1 459 ACETYL TRANSFERASE.

FT 470 858 ENOYL REDUCTASE.

FT DOMAIN 1155 1644 DEHYDRATASE.

FT DOMAIN 1645 2073 MALONYL/PALMITOYL TRANSFERASE.

FT ACT_SITE 270 270 ACETYL TRANSFERASE (BY SIMILARITY).

FT ACT_SITE 1361 1361 DEHYDRATASE (POTENTIAL).

FT ACT_SITE 1828 1828 MALONYLTRANSFERASE (BY SIMILARITY).

FT ACT_SITE 222 222 P -> R (IN REF. 1).

FT CONFLICT 222 222

FT SEQUENCE 2073 AA; 230559 MW; D928270838E7C7C4 CRC64;

alignment_scores:

Quality: 103.50 Length: 464

Ratio: 0.498 Gaps: 25

Percent Similarity: 44.828 Percent Identity: 20.690

alignment block:

US-09-303-518D-127 x FAS1_SCHPO ..

Align seg 1/1 to: FAS1_SCHPO from: 1 to: 2073

217 TCAGGCAAAATCGCCGATCATCGCGGCAAAAGCGGTACTTCAGTC 266

948 ThrGlyLeu.....MetArgGlyIleGluGluArgPheThrGlns 962

267 GGTG.....GTGATTGCCGTGAAGCAACGCAATC..... 300

962 PValGlyLysThrThrLeuIleGluAsnPhelGluAsnAspProT 979

301GAGTTGCAACGCTACGCGCCGCAAGCGTTGCGAACC 336

979 yrProValAlaAlaArgPheLeuAspAlaTyrProGluAla..... 992

337 TTAAGCGCGANGAANTNNGNCATCGATCGAATCCGGTTTGAC 386

993SerThrGlnAspLeuAsnThrGlnAspAlaGlnPheTyr 1006

387 TGCGTCGCTGATNCCGTCGCTGACGAAA.....ATCCGTCGG 424

1006 rSerLeuCySerAsnProheGlnLysProValProPheIleProAla 1023

425 TCGAT..... 432

1023 LeAspAspThrPheGluPheTyrPheLysLysAspSerLeuTrpGlnSer 1039

433 GAGCGCTTCGCCATCTTCGCAAT.....GCGATGCA 464

1040 GlnAspLeuAlaValAlaGlyGluAspValGlyArgValAlaIle 1056

465 CACCAATCCGTCGCGGACACCGCTGCTGTGATCAAGAACGCGNGC 514

1056 uGlnGlyProMetAlaAlaLysHisSerThrLysValAsnGluProAla 1073

515 ANGATTTCAAGCANGTNTGCTGTATTGACCGGTTGACGCGGTTAA 564

1073 yGlu.....LeuLeuAspGlyIleAsnGluTrpHis 1083

565 ATC...CATGTGTAGGCA.....GTGGCGGACAC.....GTGC 599

1084 IleGlnIlePheIleLysLysPheTyrAlaGlyAspLysLysIleP 1100

600 GTCTGAATGCTGCCAATCGAATCGAATGCAATTCGCGCGCGCATC 649

1100 Ile.....ValGluTyrPheGlyGlyValProP 1110

650 CGGCGGTTTGAGT..... 663

1110 roValAsnValSerHisLysSerLeuGlnSerValSerValThrGluGlu 1126

664 ...GGCAGCAGCATTCATTCATTGAGCCGCGGTGCGTCAACAAACCGT 710

1127 AlaGlySerLysValTyrLysLeuProGluIleGlySerAsnSerAla 1143

711 T..... 711

```

1143 uProserLysLysLeuTrpPheGluLeuAlaGlyProGluTrpThr 1160
712 .....TGACACATC 720
1160 rPheArgAlaIlePheThrTrpGlnArgValAlaLysGlyTrpLysLeu 1176
721 AATTATCAAGATGTAATTGCCATCGACGTTGTTTGCACAGCCGCTCT 770
1177 GlnHisAsnProVal.....ArgArgIlePheAlaPro.....Ar 1188
771 GAACACCGAGCCGTCGATTGCTTGGGTCGTCCTCAAGTCAACAACACAC 820
1188 gTyrgLysGlnArgAlaValAlaValLysGlyLysAspAsnSprThrValaIg 1205
821 GCCTCTTGGGTAC.....GTTTGGTGGCGAAGATATG 855
1205 LuLeuTyrgLysGlnArgInseryLysAsnTyrgValLeuAlaAlaArgLeuSer 1221
856 CAATTTACTGCGGCGCAATTGGTTGACGCGACAGCAACCCGCGATTTCGG 905
1222 .....TyrAspGlyGluThrIleValIleValSerMe 1231
906 TTCGGATTGACGCGCGGATTACACAGCGCGCAC...GATTATTGG 952
1231 tPheGluAsnArgAsnAlaLeuLysLysGluValHisLeuAspPheLeuP 1248
953 GACGCTAC.....CACAAATCAGATTTCGGTATCGAAGAA 987
1248 helTygTygGluProSerAlaGlyTyrSerProValSerGluLeuAsp 1264
988 GGCCCGACGCAAGAGCTG.....TTCGGCTGGGTT..... 1017
1265 GlyArgAsnAspArgIleLysHisPheTyrgTrpAlaLeuTrpPheGlyG 1281
1018 .GCGCGCGACCGGACAAATACTCATACGCGCTACGACCCCTCGGCCATT 1066
1281 uGluProTyrgProGluAsnAlaSerIleThrAspThrPheThrGlyProg 1298
1067 TCCTGAAA.....AACAACTCTTCAGTTCACGACGACGCGTCAC 1107
1298 LuValThrValThrGlyAsnMetIleGluAspPheCysArgTrpValGly 1314
1108 GGTGGCGACCGCGCCATGTCGTCGATTCACGACGCGGATATGCC 1157
1315 AsnHisAsnGluAlaLysTrpHisArgAlaIleArgLysArgMetAlaPr 1331
1158 GCTAGAC.....ATCC 1168
1331 oMetAspPheAlaIleValValGlyTrpGlnAlaIleThrLysAlaIleP 1348
1169 TGCCCTACCCCTGTTTGGCGATTTA.....ATCGTCGCGATACCGAC 1212
1348 heProLysAlaIleAspGlyAspLeuAlaGlyLeuValHisLeuSerSn 1364
1213 AGCGCGCAAGCATGGTGGTTCGTAATTGGACGACGACAGAC 1254
1365 SerPheArgMetValGlySerHisSerLeuMetGluGlyAsp 1378
seq_name: SwissProt_40:GID_BACSU
seq_documentation_block:
ID GID_BACSU STANDARD; PRT; 435 AA.
AC P39815;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein gId.
GN GID.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-168;
RA Foulger D., Errington J.;
RT "Cloning and sequencing of 7.5Kbp of DNA from Bacillus subtilis upstream
RT of the codv gene."
RL Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-363 FROM N.A.
RC STRAIN-168 / 865;
RA de Jong S.;
RT "Cloning and sequencing of the top1 gene, the gene encoding B.
RT subtilis DNA topoisomerase I."
RL Submitted (XXX-1994) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE GIDA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AJ000975; CA04423.1; -.
CC EMBL: Z99112; CAB3486.1; -.
CC EMBL: I27797; AAA22764.1; -.
CC Subtilisin; B01.008; gId.
CC InterPro: IPR002218; GIDA.
CC InterPro: IPR000205; NAD-binding.
CC ProDom: PD003738; GIDA; 1.
CC PROSITE: PS01280; GIDA.1; FALSE_NEG.
CC PROSITE: PS01281; GIDA.2; 1.
CC Complete proteome.
CC SEQUENCE 435 AA; 48063 MW; 6C7AB028F484B683 CRC64;

```

```

alignment_scores:
  Quality: 103.00      Length: 295
  Ratio: 0.824         Gaps: 11
  Percent Similarity: 42.373   Percent Identity: 20.678

```

alignment_block:

US-09-303-518D-127 x GID_BACSU ..

Align seg 1/1 to: GID_BACSU from: 1 to: 435

```

331 GCAAACTTAAGCGCGGANGAANTNNGNNGCAATGATCCAAATCCGTTT 380
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
11 AlaGlyLeuAlaGlySerGluAlaIleTrpGlnLeuAlaLysArgGly 27
381 GTGACACGCGCTG...CGTANCCGTCGCTTCAGCAAAATCCGCGCTGG 427
: : : : : : : : : : : : : : : : : : : : : : : : : : :
27 eGlnValLysLeuTyrgGluMetArgProValLysGlnThrProAlaHis 44
428 ATCCGAGCCGCTGCGCATCTGCTG.....AATCGATGGACACCAAT 471
: : : : : : : : : : : : : : : : : : : : : : : : : : :
44 tThrAspLysPheAlaGluLeuValCysSerAsnSerLeuAlaGlySerAsn 60
472 CGCTNGCGGCGACCCCTGTGGTGTGATCAAGAAGCCGCGANGATTT 521
||| |||:||||:||||:||||:||||:||||:||||:||||:||||:
61 ThrLeu...AlaAsnAlaValGlyValLeuLysGluGluMet..... 73
522 CAACAGCAGTTCCTGCTGATGAGCCGTTTGACCGGACGCTAAATTCATG 571
74 .....ArgAlaLeuAspS 78
572 TGTGTAAGCAGCTGCGCGAGAC.....GTGCGCTGTGAAAATGCT 612
: : : : : : : : : : : : : : : : : : : : : : : : : : :
78 eValIleIleLeuAlaIleAlaAspGlyCysSerValProAlaGlyGlyAla 94
613 GCCAATCATGCAACACATGTAATTGGCGGCGCG..... 645
: : : : : : : : : : : : : : : : : : : : : : : : : : :
95 LeuAlaValAspArgHisGluPheAlaIleSerValThrAsnArgVal 111

```



```

344 GCGAAGCAANTNNGNCATCTGATCCATCCGTTTGGACGTCG 393
130 sprenclunhrlysprolaialepnehlthrcllylelysalvalasp 146
394 CGTACCCGTCGTCAGCAAAATCCCTGCGTCGATGCCGACCGCTTGC 443
147 LeuleuAlaProlyrArgArg.....GlyGlyLysIleG1 158
444 CATTCTCGCATGGATGGACACCAATCCGTCNGCGCAGACCTTGC 493
158 yLeuphecllyGlyAla.....GlyValGlyLysThrV 169
494 TTGTGATCAAGAGCCGNCANGATTTTCAGACGANGTGTGCTATTG 543
169 alLeuilemetcluleuileasnailleAlaLys..... 180
544 ACCCGTTTGACCGACGCGTAAATCCATGCTGTGAAGCAGCTGGCGAGA 593
181 .....AlaHisGlyGlyValSerValPheGlyGlyValGlyLuar 194
594 CGTGGCGTGTGAAATGCTGGCAACATGAAACACATGATTCGGCGGC 643
194 gThrArgGlyGlyAsnAspLeuTyMetclumeltygluSerGlyVal 211
644 CGCATCCGCGCGTTGAGTGACGACGACATTCATTTCATT..... 684
211 leasnGluserAsnleuSerGlySerValAlaLeuValtyGlyGln 227
685 .....GAGCCGCTGCTGCAACAAACCCGTTGG..... 714
228 MetasnGluProGlyAlaArgMetArgValGlyLeuThrAlaLeuTh 244
745 .....ACATCAATTATCAAGTATTAATGTC...A 742
244 rMetclagluTyrrheArgaspIleasnArgIlnspalleuLeuphei 261
743 TCGGACGTTTGTTCACACAGCCGCTGACACCGCGGCTGATTCCT 792
261 leasnIlnlephearPheValGlnAlaGlySerGluValSerAlaLeu 277
793 TTGGGT.....GGTTCACAGTCAACAAACCCGCTTGGCTGACCGT 836
278 leuGlyArgMetProSerAlaValGlyTyrlnProThrAlaGlyThG1 294
837 TTGGGTGCGAAGATATGCAATACGCGGCGAATTGTTGACGAG 886
294 uMetclYAlaLeuGlnGlnAlaArgIleThr..... 303
887 ACAACCGCGTATTCGCGTTGCGTATTCGAGCGCGATTCACAAAGC 936
304 .....SerThrThrGlnGlySerIleThrSerIle 313
937 GCGCAGATTAATTGGACGCTACCAACATGATTCGTTATCGAAGA 986
314 GlnAlaValTyVal..... 318
987 AGGCGCGCAAGAGCTGTCGCGTGGTTCGCGCGCGGCAAT 1036
319 .....ProAlaAspAspLeu.....ThrAspProAlaPro..... 328
1037 ACTCCATCAGCGTACGACCTCGGCAATTCGTAACAAACACTGCTC 1086
329 .....AlaThrThrPheAlaHis.....Leu 335
1087 AAGTTCAGACGCGCTCAAGCGTGGCGACGCGCGCATGGTCCGATTGG 1136
336 AspAlaThrThrValLeuSer.....ArgGlyLeuAlaAlaLysG1 349
1137 TACTTACGAGCGCGTAATGCGCTAGACATCTGCTGCTGCTTGGC 1186
349 yIleTyrrProAlaValAspProLeuAspSerThrSerThr...MetLeuG 365

```

```

1187 GCGATTAATATCGCGCATAC.....GACAGCGCGCAA..... 1221
365 lnProvalIleValGlySerGlnHisTyrrAspThrAlaGlnLeuValys 361
1222 .....GCATTGGG 1229
382 LysThrLeuGlnArgTyrrGlyLeuGlnAspIleIleAlaLeuGln 398
1230 TTGCTTGGAATTGGACGACAGACAGAC..... 1254
398 yIleAspGluLeuSerGlnGlnAspArgLeuValAlaAspArgAlaArgL 415
1255 .....CTCGCTTGTGACGCTGCTGTC..... 1278
415 yslleGluArgPheLeuSerGlnProPhePheValAlaGluValPheThr 431
1279 .....CCGGGCAAAATAC 1290
432 GlySerProGlyLysTyrr 437
seq_name: SwissProt_40:RNFC_ECO57
seq_documentation_block:
ID RNFC_ECO57 STANDARD; PRT; 740 AA.
AC P58324;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE Electron transport complex protein rnfc.
GN RNFC OR Z2636 OR ECS2338.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., Iit, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postal G., Hackett J., Klink S., Boulton A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoustis K.,
RA Apodaca J.A., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shimagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: May be part of a membrane complex involved in electron
CC transport (By similarity).
CC -!- COFACTOR: Binds 2 4FE-4S clusters (potential).
CC -!- SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
CC rnfD, rnfE and rnfG (By similarity).
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -!- SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC RNFC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR ENBL: AF005386; BAG56618.1: -
DR ENBL: AP002558; BAB35761.1: -
DR PROSITE: PS00196; 4FE4S_FERREDOXIN; 2.
FW Electron transport; Iron-sulfur; 4Fe-4S; Inner membrane.
FT METAL 377 377 IRON-SULFUR (4FE-4S) (POTENTIAL);
FT METAL 380 380 IRON-SULFUR (4FE-4S) (POTENTIAL);
FT METAL 383 383 IRON-SULFUR (4FE-4S) (POTENTIAL);
FT METAL 387 387 IRON-SULFUR (4FE-4S) (POTENTIAL);
FT METAL 416 416 IRON-SULFUR (4FE-4S) (POTENTIAL);
FT METAL 419 419 IRON-SULFUR (4FE-4S) (POTENTIAL);
FT METAL 422 422 IRON-SULFUR (4FE-4S) (POTENTIAL);
FT METAL 426 426 IRON-SULFUR (4FE-4S) (POTENTIAL);
FT CONFLICT 651 651 R -> RKAVEAIAIARAKRLPQQAANEPEEPVDP
      (IN REF. 2)
SO SEQUENCE 740 AA; 80297 MW; 8EB8A9E22DEC56F8 CRC64;

alignment_scores:
  Quality: 103.00      Length: 324
  Ratio: 0.665         Gaps: 13
  Percent Similarity: 47.840   Percent Identity: 21.914

alignment_block:
  US-09-303-518d-127 x RNFC_ECO57
  Align seg 1/1 to: RNFC_ECO57 from: 1 to: 740

100 GAAGATATGCGGATGCGCCCTGATGAAATGCAAGGAGGCGATC 149
    :::::  |||  ::  ::::  |||||
48 LysGlnHisIleGlyAlaGlnIleuGlyValSerValGlyAsp 64
150 CGTCAAAAGGCGCAAGTGTCTT.....GAAGACAAAGNATCGCG 193
    |||  ::::  |||||  |||  |||
64 sValLeuArgGlyGlnProLeuThrArgGlyArgGlyLysMetLeuPro 80
194 GCGTGTGTTTACCGCGCCGCTTCAGGCAAAATCGCGCGCATC.... 237
    |||||  ::::  |||||  |||||  |||||
81 .....ValHisAlaProThrSerGlyThrValThrAlaIleLeuPro 94
238 .....CATCGCGCGCAAAAGCGCGTACTGACGCGCGTGTAT 275
    |||  ::  |||  |||||  ::::  |||
95 HisSerThrAlaHisProSerAlaLeuAlaGluLeu...SerValIle1 110
276 TCGCGTGAAGGCAACGAC..... 294
110 eAspAlaAspGlyGlnAspCysTrpIleProArgAspGlyTrpAlaAsp 127
295 .....GAATCGAGTTGCAAGCGCTAGCGCGCGCGAGCG 327
    |||  ::::  |||||  ::::  |||
127 yArgSerArgArgArgGlnGluIleuIleGluArgIleHisGlnPheGly 143
328 TTGGCAACTTGAAGCGCGGAGCAAAATNNGNCATTCGATCCGG 377
    ::::  |||||  ::::  |||||  ::::  |||
144 ValAlaIleGlyLeuGlyLys.....AlaGly 151
378 TTGTGAGATGCGCGCGGTANCGCTCCGTTGACAAATCCCTGCGCG 427
    |||  ::::  |||||  ::::  |||
151 yPheProThrGlyVal.....LysLeuGlnGlyGly 162
428 ATGCCAGCGCGTTCGCGCATTTGTCATGATGATGATGATGATGATG 477
    ::::  |||||  ::::  |||||  ::::  |||
162 lAspArgIleGlnThrIleuIleIleAsnAlaIleGlnGlnGlnGln 178
478 GCGGCGAGCCCTGTGTGTGATCAAGAGAGCGCGCGAGATTCAGAG 527
    ::::  |||||  ::::  |||||  ::::  |||
179 lIleThrAlaAspAspArgLeuMetGlnAspCysAlaAlaGlnVal 195
528 ANGTNTGATGATGATGATGATGATGATGATGATGATGATGATGAT 558
    ::::  |||||  ::::  |||||  ::::  |||
195 uGlyIleArgIleLeuAlaHisIleGlnIleProArgGlnIleLeu 212
559 .....CGTAAATTCATGTGTGTAAGGACGCTGCG 588

```

```

212 lylIleGlnAspAsnLysProGlnAlaIleSerMetLeuArgAlaValLeu 228
589 GCGAGCGTCCGCGCTGTAATAATGCTGCCAACATCGAAACATGATTCG 638
    |||||  ::::  |||||  ::::  |||||
229 AlaAsp.....SerHisAspIleSerMetArgValIlePr 240
639 CGGCGCCCATCGCGCGCT.....TTGAGTG 664
240 oThrLysTrpProSerGlyGlyAlaLysGlnLeuThrTrpIleLeuThr 257
665 GCAAGCGATTCATTTTCATGACCGCGGTGCGCAAAACCGTTGG 714
    ||  ::::  |||  |||  |||  |||
257 lLysGlnVal.....ProHisGlyGlyArgSerSerAla 269
715 ACCATCATATATCAAGATGA.....ATGCGATCGAGCGTGTGT 755
    ::  ::::  |||||  |||||  |||||
270 GlyValLeuMetGlnAsnValGlyThrAlaTrpAlaValLysArgAla 286
756 TGCAACAGCGCGCTGTGACACAGCGCGGTGATGCTGGTGGTTC 805
    |||||  ::::  |||||  |||||  |||||
286 lIleAspGlyGlnProIleThrGlnArgValValThrLeuThrGlyLys 303
806 AAGTCAACAAACACGCGCTTGTGCGTACGTTGGTGGCGAAAGATCG 855
    ::  ::::  |||  ::  ::  ::::  |||
303 lAlleAlaArgProGlyAsnValTrpAlaArgLeuGlyThrProValArg 319
856 CAATTT.....ACTGGGCGCAATTTGTTGACGACACACCGCGTAT 899
320 HisLeuLeuAsnAspAlaGlyPheCysProSerAlaAspIleMetVal 336
900 TTCCGTTGCGTATTTGAACGCG 921
    |||||  ::::  |||  |||
336 eMetGlyGlyProLeuMetGly 343

seq_name: SwissProt_40:RNFC_ECOLI

seq_documentation_block:
ID RNFC_ECOLI STANDARD; PRT; 740 AA.
AC P77611:
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Electron transport complex protein rnfc.
GN RNFC OR B1629.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.C., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).

[2]
RN SEQUENCE FROM N.A.
RX STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nasimoto H., Nishio Y., Oshima T., Saito N.,
RA Sampaio K., Seki Y., Sivasubramanian S., Tagami H., Takeuchi T.,
RA Takeuchi K., Takeuchi Y., Wada C., Yamamoto Y., Horikuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
-i- FUNCTION: May be part of a membrane complex involved in electron

```

```

CC transport (By similarity).
CC -1- COFACTOR: Binds 2 4Fe-4S clusters (Potential).
CC -1- SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
CC rnfD, rnfE and rnfG (By similarity).
CC -1- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -1- SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC RNF SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000258; AAC74701.1; -.
DR EMBL: D90806; BAA15384.1; -.
DR EMBL: D90807; BAA15391.1; -.
DR EMBL: D90808; BAA15414.1; -.
DR HSSP: P00198; 1FCA.
DR EcoGene: EG13935; rnfC.
DR InterPro: IPR001450; 4Fe4S-ferredoxin.
DR InterPro: IPR001949; Complex1_51K.
DR Pfam: PF01512; Complex1_51K; 1.
DR PROSITE: PS00198; 4FE4S_FERREDOXIN; 2.
KW Electron transport; Iron-sulfur; 4Fe-4S; Inner membrane.
KW Complete proteome.
FT METAL 377 377 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 380 380 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 383 383 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 387 387 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 416 416 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 419 419 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 422 422 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 426 426 IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ SEQUENCE 740 AA; 80171 MW; 381506475CD6E01F CRC64;

alignment_scores:
Quality: 103.00 Length: 324
Ratio: 0.665 Gaps: 13
Percent Similarity: 47.840 Percent Identity: 21.914

alignment_block:
US-09-303-518D-127 x RNF_CECOLI ..
Align seg 1/1 to: RNF_CECOLI from: 1 to: 740

100 GAAGATATGCGGATGCGCCCTGATGAAGTCAAGGAGGCGATGC 149
   ::::: ||| ::::: ||| ::::: ||| ::::: |||
48 LysGlnHisIleGlyAlaGluGlyIleuGlyValSerValGlyAspIy 64
150 CGTCAAAAAGGCCAAGTGTGTTT.....GAAGCAAAAAGNATCCGG 193
   ||| ::||| ||| ||| ::||| ||| |||
64 sValIleuArgGlyIleProIleuThrArgGlyArgGlyLysMetIleuPro. 80
194 GCGTGTGTTTACGCGCCGTTTCAGGCAAAATCGCCGATC..... 237
   ||||| ::||| ::||| ::||| ::||| ::|||
81 .....ValHisAlaProThrSerGlyThrValThrAlaIleAlaPro 94
238 .....CATCGCGGCAAAAGCGCGTACTTCAGTGGTGTGAT 275
   ||| ::| ||| ||| ||| ::||| |||
95 HisSerThrAlaHisProSerAlaIleuAlaGluIleu...SerValIleI 110
276 TGCCGTTGAAGCAACGAC..... 294
   | ::::: ||| ::::: |||
110 eAspIaAspGlyIleAspCysTrpIleProArgAspGlyTrpIaAspTr 127
295 .....GAATCGAGTTCGAAGCGTACGCGCCGCAAGCG 327
   ||| ::||| ||| ::||| |||
127 yTrArgThrArgSerArgGluIleuIleGlyIleHisGlnPheGly 143

```

```

328 TTGGCAACTTAAGCGCGGANGAANTNNGNCATCTGATCCATCCGG 377
   ::||| ::||| ::||| ::|||
144 ValAlaIleuGlyGly.....AlaG 151
378 TTGTGAGTCGCGTGGGTANCCGTCGTCAGCAAAATCCGTCCGTGG 427
   ||| ::||| ::||| ::|||
151 yPheProThrGlyAla.....LysIleGlnGlyG 162
428 ATGGCGCGCGTTCGCCATCTTCATGCAATGCAACCAATCCGCTN 477
   ::||| ::||| ::||| ::|||
162 LysAspIleGluThrIleuIleIleAsnAlaIleGlyGlyProTyr 178
478 GCGGCAACCCCTGTGTGATGATCAAGAAGCCGANGANTTTCAGAG 527
   ::||| ::||| ::||| ::|||
179 IleThrAlaAspAspArgIleuMetGlnAspCysAlaIleValG 195
528 ANGTNTGCTGATATGAGCCGTTGACCGAG..... 558
   ::||| ::||| ::||| ::|||
195 uGlyIleArgIleIleuAlaHisIleuGlnProArgGluIleuIleG 212
559 .....CGTAAATCCATGTGTGAAGCAGCGTGGC 588
212 LylleGluAspAsnLysProGlnAlaIleSerMetLeuArgAlaVal 228
589 GCAGAGCGTGGCGTGTGAATGCTGCCAATGCAACATCAATTCGG 638
   |||||
229 AlaAsp.....SerAsnAspIleSerLeuArgValIlePr 240
639 CGGCGCCGATCCGCGCGGT.....TTGAGTG 664
   ::||| ::||| ::||| ::|||
240 oThrLysTyrProSerGlyAlaLysGlnLeuThrTyrIleLeuThrG 257
665 GCAGCGCATTCATTCATTGACCGCGGTGCAACAACAAACCGTTGG 714
   ||| ::||| ::||| ::|||
257 LysGlnVal.....ProHisGlyIleArgSerSerAspIle 269
715 ACCATCATATCAAGATGTA.....ATTGCATCGGAGCGTGTGT 755
   ::||| ::||| ::||| ::|||
270 GlyValIleuMetGlnAsnValGlyThrAlaTyrAlaLysArgAla 286
756 TGCAACAGCGCGTGTGAACACGAGCGCGTATGCTTGGGTGTC 805
   ||||| ::||| ::||| ::|||
286 IleAspGlyIleProIleThrGluArgValThrIleThrGlyAla 303
806 AAGTCAACAAACCAACGCTCTGCTACCGTTTGGGTGCAAGATGCG 855
   ::||| ::||| ::||| ::|||
303 IleIleAlaArgProGlyAsnValTrpAlaArgIleuGlyThrProVal 319
856 CAATT.....ACTGGGCGCAATGTGTGACGACAGACAGCGCGTAT 899
   ||||| ::||| ::||| ::|||
320 HisIleuAsnAspAlaGlyPheCysProSerAlaAspIleMetVal 336
900 TTCGCGTTCGATATGAACGCG 921
   ||||| ::||| ::||| ::|||
336 eMetGlyGlyProIleuMetGly 343

seq_name: SwissProt_40: MUC1_MESAU

seq_documentation_block:
ID MUC1_MESAU STANDARD; PRT; 676 AA.
AC Q60528;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Mucin 1 precursor.
GN MUC1.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
NCBI_TaxID=10036;
RX [1]

```



```

352 F.....giuseraspleualaThrThrProIleaspserylt 365
280 CGGCATACGACGACGATGAGTACGCCCTTTGCCCGCATGATGGCG 231
365 hSerIleSerThrThrLysAlaProAlaThrThrProValHisAsnGly 381
230 GCGATTTCCTGGAACGCGCGGTA.....AACACCGCGCGG 190
382 SerLeuValProThrThrSerSerValLeuGlySerAlaThrThr.... 396
189 ATNCTTTTGTCTCAACAGACGCTGGCTTTTTCAGCGCATCGCT. 141
397 .....LeuIleHisAsnAspThrSerThrMetAlaThrThrThrPro 411
140 .....TCCTGACTTCATCAACGAGCGCGCATCCGCGCATTTCT 102
411 aGlyAsnGlyThrGlnSerSerValProSerArgHisProValThrPro 427
101 TCGCCA.....AGCAACGCGACTTCGTAATGACG 72
428 ThrProProAlaValSerSerAsnSerThrIleAlaLeuSer 441
seq.name: SwissProt_40:DAN4_YEAST

```

```

seq_documentation_block:
ID DAN4_YEAST STANDARD; PRT; 1161 AA.
AC P41719;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Cell wall protein DAN4 precursor.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_Taxid=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Scaresz T.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE-21113168; PubMed-11160904;
RA Cohen B.D., Serfl O., Abramova N.E., Davies K.J., Lowry C.V.;
RT "Induction and repression of DAN1 and the family of anaerobic
manoprotein genes in Saccharomyces cerevisiae occurs through a
RT complex array of regulatory sites";
RL Nucleic Acids Res. 29:799-808(2001).
CC -1- FUNCTION: COMPONENT OF THE CELL WALL (By similarity).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
(Potential).
CC -1- PPM: EXTENSIVELY O-GLYCOSYLATED (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SRP1 / TPI1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z49651; CAA89684.1; -
DR SGD: S0003912; DAN4.
DR InterPro: IPR000992; SRP1_TPI1.
DR Pfam: PF00660; SRP1_TPI1.1.
DR PROSITE: PS00724; SRP1_TPI1.1.
KW Cell wall; Glycoprotein; Membrane; GPI-anchor; Signal.
FT STGNL 1 24
FT CHAIN 25 1146 CELL WALL PROTEIN DAN4.
FT PROPEP 1147 1161 REMOVED IN MATURE FORM (POTENTIAL).
FT LIPID 1146 1146 GPI-ANCHOR (POTENTIAL).

```

```

SQ SEQUENCE 1161 AA; 118358 MW; 7954C15D69F0CA58 CRC64;
alignment_scores:
Quality: 101.00 Length: 451
Ratio: 0.459 Gaps: 18
Percent Similarity: 48.780 Percent Identity: 21.508
alignment_block:
US-09-303-518d-127/rev x DAN4_YEAST ..
Align seg 1/1 to: DAN4_YEAST from: 1 to: 1161
1337 TCCTTCACANNGTTTCACACCTTACGAAACGCGGCNAT..... 1293
346 AlaThrThrSerAlaThrThrThrThrThrSerAspThrThrIle 362
1292 .TCGTATTCCTCCGCGGACGACGAGTCACAAAGCAGAGCTTCCTTCGT 1244
362 rSerSerSerProSerGlnVal.....ThrSerSer 373
1243 CCAATTCCAAAGCAACCAATGCTGCGCGCTGCGGTATCGCGACGATT 1194
373 IagIuProThrThrValSerGluValThrSerSerValGluProThr 388
1193 AATTCGCGCAAAAGCAGGTAAGGACGAGATGCTAGGCGCATTCGCGCTC 1144
389 .....ArgSerSerGlnValThrSerSerIagIuProThr 400
1143 GTAAGTACCAATGCGACATGCGCGGTCGCCACCGTTGACGCGTCTG 1094
400 rThrValSerGluPheThrSerSerValGluProThrArgSerSerGln 417
1093 TGAACCTTGAAGAGTTGTTTTCAGGAATAGCGGAGGTCGACCGTG 1044
417 alThrSerSerAla.....GluProThrThrVal 426
1043 ATGAGATATTTCGCGGCTGCGCGCAACCCGACGACGACGCTTCGCT 994
427 SerGluPheThrSerSerValGluProThr..... 436
993 GCGGCTTCCTTCGATACGGAATCTGATGTGTAGCGTCCCAATTAAT 944
437 ArgSerSerGlnValThrSer.....S 444
943 CGTGGCGGCTTGTGTAATCGCGCTCAATACCAACCGAATACG 894
444 eAlaGluProThrThrValSerGluPheThrSerSerValGluProThr 460
893 CGGTGTCGTCGTCACCAATTCG.....CCGCGATTAATTGGCAT.. 852
461 ArgSerSerGlnValThrSerSerAlaGluProThrThrValSerGlu 477
851 .ACTTCGACCCCAACGATGACGACGACGCTGTTGTCATCTGAG 803
477 eThrSerSerValGluProThrThrSerSerGlnValThrSerSerAla 494
802 AACCAACCAACCAATCAGCGCGCTGCTTCAAGACGCGCTGTGCAAC 753
494 IuPro.....ThrThrValSerGluPheThrSerSerValGluPro 507
752 AAGCGTCGATGCGCATTTACATCTGATAATGATGTCGCAACGGTTT 703
508 IlaArgSerSerGlnValThrSer.....SerAlaGluProThrThrVa 522
702 GTTTCACGACGCGCTCAATGAATGATGTCGACCTCAACACCG 653
522 IserGluValThrSerSerValGlu.....ProIleArgSerS 535
652 CCGGATGCGGCGCGCAATTCATGTTTCATGTTGCGCATTTTCA 603
535 eGlnValThrThrThrGluProValSerSerPheGlyThrThrPheSer 551

```



```

104 raaatgagatgtrpserargleuargleuval..... 116
105 .....
656 CCGGCGGATGCGGCGCGGATTCATG.....TGTTCGATGT 616
117 ..AspIysleuValleuGluTyrIleuSerPheTyrIleuAspTyr 132
615 GCGACGATTTTCAGCGGCGCGCTCGGCC..... 586
133 ArgIeuValValProLysHisIlePheAlaArgIerSerIeuValG 149
585 ..ACGTGCTTACACACATGATTTTACGCTCGGTCAGAGCGCTCAATACC 537
149 yGluCysCysAlaHis.....AlaThrAlaArgTyrA 160
536 AGCAACNCTGCTGAATCCTGCGG..... 510
160 speIthrasProAsnValAlaSerPheGlySerValAlaGluTyr 176
509 ..GCTTCTTGATCACAACGAGGCTCGCGCGGATTCGTCGA 461
176 pAlSerIeuPheSerAsnValGInSerIeuAspAspIeuPheSer 193
460 TCACA...TTGACGAAGATGCGCAACGCTCGGATCGACGCGGAGATT 414
193 IeAlaCysIeuThrLysIleAlaHisArgIerMetAsnAlaAla 209
413 TTGCTGACGAGGAGTACGACGCGGATTCGCAACGCGGATTCGATCG 364
210 LeuLysAsnGlnAlaThrArgLysProSerSerPro.....G 223
363 ATTCGNNCANNANTCTGCGCGGCTGAAGTTTGCACAGCTCGGCGGCT 314
223 nThrIrgInAspSerProle..... 231
313 AGCGTTGCACTGATTTGCTGCTGCTTCAACGCAATCAACGACGCG 264
232 .....LeuThrMetAlaProSerThrProVal..... 240
263 TGAAGACGCGGCTTTCGCGCGGATGATGCGGCGGATTTGCT 219
241 ...SerValGlySerThrProProSerThrProSerValIeuProLeu 256
218 ....GAACGCGCGGTAACACGCGCGGATTCCTTTGCTTCA 173
256 aLysGlnIeuAlaProMetAsnValCysLysAlaHisIleGlnAlaSer 273
172 ACAGCACTGGGCTTTTACGCGCATCGCT..... 141
273 sngInserArgThrIeuThrAlaSerProProGluGlnIleProLeu 289
140 .....TCCTGACTTTCATCNCGCGCGGATTCGATACC 112
290 MetGluProGlnValIeuValAsnProGlnValIeuProGluArgL 306
111 GGCATATTCCTGCGCAACGCGCATTCGTAATGACGCGCGGCTCAT 62
306 rSerIeuSerLysPro.....ValSerIeuProProThrProSer 319
61 AATGACTTGCTCGGCTCGCGCGGATTCGATGAGCT 18
320 .....SerProLysValGlyValIeuTyrPro 328
seq_name: swissprot_40:RNFC_BUCAI
seq_documentation_block:
ID RNFC_BUCAI STANDARD; PRT; 473 AA.
AC P57215;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Electron transport complex protein rnfc.
GN RNFC OR BU115.

```

```

OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TOKYO 1998;
RX MEDLINE=20445173; PubMed=10993077;
RT Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS."
RL Nature 407:81-86(2000).
CC -1- FUNCTION: May be part of a membrane complex involved in electron
CC transport (By similarity).
CC -1- COFACTOR: Binds 2 4Fe-4S clusters (Potential).
CC -1- SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
CC rnfD, rnfE and rnfG (By similarity).
CC -1- SUBCELLULAR LOCATION: Membrane-associated (By similarity).
CC -1- SIMILARITY: BELONGS TO THE 4FEAS BACTERIAL-TYPE FERREDOXIN FAMILY.
CC RNFC SUBFAMILY.

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to license@isb-sdb.ch).

```

DR EMBL; AP001118; BAB12833.1;
DR InterPro; IPR001450; 4Fe4S_ferredoxin.
DR InterPro; IPR001949; Complex1_51K.
DR Pfam; PF00057; Fer4; 2.
DR Pfam; PF00152; Complex1_51K; 1.
DR PROSITE; PS00198; 4FEAS_FERREDOXIN; 2.
KW Electron transport; Iron-sulfur; 4Fe-4S; Complete proteome.
FT METAL 337 337 IRON-SULFUR (4Fe-4S) (POTENTIAL).
FT METAL 340 340 IRON-SULFUR (4Fe-4S) (POTENTIAL).
FT METAL 343 343 IRON-SULFUR (4Fe-4S) (POTENTIAL).
FT METAL 347 347 IRON-SULFUR (4Fe-4S) (POTENTIAL).
FT METAL 376 376 IRON-SULFUR (4Fe-4S) (POTENTIAL).
FT METAL 379 379 IRON-SULFUR (4Fe-4S) (POTENTIAL).
FT METAL 382 382 IRON-SULFUR (4Fe-4S) (POTENTIAL).
FT METAL 386 386 IRON-SULFUR (4Fe-4S) (POTENTIAL).
SQ SEQUENCE 473 AA; 54563 MW; 04B64102F1313AE6 CRC64;

```

alignment_scores:
Quality: 100.50 Length: 462
Ratio: 0.455 Gaps: 22
Percent Similarity: 47.835 Percent Identity: 19.697

alignment_block:

US-09-303-518d-127 x RNFC_BUCAI ..
Align seg 1/1 to: RNFC_BUCAI from: 1 to: 473

```

100 GAAGAATATGCGCGGATATGCGCGGCTTCGTAATGACGAGGCGATGC 149
13 AspspIYrHIsAsnValLys.....LeuArgValLysIleAsnGlu 27
150 CGTCAAAAAGGCCAA...GTGCTGTTGAAGACCAAAAGNATCCGGGCG 196
27 sValIeuArgGlyGlnProIeuIlePheSerAspPhe.....AsnY 42
197 TGTGTTTACCGCGCCNGTTTCAGCAAAATCGCGGATTCATCGGCGC 246
42 alProValHIsAlaProThrSerCylIeuIleGlnsIleCysPheAsn 58
247 GAAGAAGCGGACTTCAGTGGTGGTGGATTCGCGGATTCGCGGATTC 288
59 SerAspSerIleLysLysAsnIleLysIleValIleSerProAspTyrIle 75

```

[illegible]

```

1      |||          :||           ..::.....serGluYstFtnCysIIleargCySg 341
2      ulYsaSnInLuSerIle.....
3
4      1136 GTACTTACAGCGCGTAATGCCCTAGACAATTCGTACCCTGCCTTT... 1182
5              ::||   |||:::|||||    |||
6      341 lYTtCYseSTeTyRvAlcYSPrOVAlSaSnInLeubPrgOnIngInLueUTyr 357
7
8      1183 .....TTGGCGGAATTAAATCGTGCCGGATACAGCAACGCGGACAACTT 1226
9
10     358 TrPYrIlIelYsaSnlYsaSnHIsValGlThILySLySHIStyrVAlle 374
11         | |||:::|||
12     374 uAsPCyStllEGU.....cYslYsalAcYSGnlLYsValIC 386
13         |||:::|
14     1277 GCCCGGGGC.....AAtATAC.....GAATNGCCCCG 1302
15         |||:::|
16     386 ySPROSeRTyrlEProLeUVallYstYrPhElleGIingInLUlysaSnIle 402
17         |||:::|
18     1303 CTGTTCGTAAGTCGTGGAAACCTTGAGAGAAGAA 1338
19         |||:::|
20     403 LeUlYsaSnIlIErhLeGLuIasASnaGrLyLS 414
21
seq_name: SwissProt_40:HFC1_HUMAN
seq_documentation_block:
ID       HFC1_HUMAN                STANDARD;             PRT; 2035 AA.
AC        PS1610;
DT       01-OCT-1996 (Rel. 34, Created)
DR       01-OCT-1996 (Rel. 34, last sequence update)
DS       16-OCT-2001 (Rel. 40, last annotation update)
DE       "The VP16 accessory protein HCF is a family of polypeptides processed from a large precursor protein." ;
GN       HCF1 OR HCF1.
GS       Homo sapiens (human).
OC       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX       NCBI_TaxId=9606;
[1]
RN       [1]
RP       SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX       MEDLINE=93327419; PubMed=8392914;
RA       Wilson A.C., Lamarco K., Peterson M.G., Herr W.;
RT       "the VP16 accessory protein HCF is a family of polypeptides processed from a large precursor protein.";
RL       Cell 74:115-125(1993).
[2]
RN       [2]
RP       SEQUENCE OF 65-2035 FROM N.A.
RC       TISSUE=Fetal brain;
RX       MEDLINE=95130107; PubMed=7829097;
RA       Fratlini A., Faramda S., Redolfi E., Zucchi I., Villa A.,
RT       Patrosso M.C., Strina D., Susani L., Vezoni P.;
RL       "Genomic organization of the human VP16 accessory protein, a housekeeping gene (HCF1) mapping to xq28.";
RM       Genomics 23:30-35(1994).
[3]
RN       [3]
RP       PROCESSING.
RX       MEDLINE=96033796; PubMed=7590226;
RA       Wilson A.C., Peterson M.G., Herr W.;
RT       "the HCF repeat is an unusual proteolytic cleavage signal.";
RL       Gene Dev. 9:2445-2458(1995).
[4]
RN       [4]
RP       FUNCTION: UPON LYRIC INFECTION OF PERMISSIVE CELLS, THE HSV TRANSCRIPTIVATOR PROTEIN VP16 ASSOCIATES WITH HCM. BINDING TO HCM ACTIVATES VP16 FOR ASSOCIATION OF THE HSV IMMEDIATE EARLY GENES. ACTIVING TRANSCRIPTION OF THE HSV DNA COMPLEX RESPONSIBLE FOR PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX EARLY GENES. SUBUNT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM 110 TO 150 kDa AND A MINOR 300 kDa POLYPEPTIDE. THE MAJORITY OF NONCOVALENTLY ASSOCIATED.
[5]
RN       [5]
RP       SOBELCELLULAR LOCATION: Nuclear.
RX       -| ALTERNATIVE PRODUCTS: A SHORT FORM MAY ARISE BY ALTERNATIVE SPLICING. IN THIS VARIANT THE N- AND THE C-TERMINAL FRAGMENTS FAIL TO ASSOCIATE.
```



```

55 leaIaValIleasnIleLysAlaGlyGlnIleLysValAlaGlyAspVal 71
271 GTGATTCGCGTTGAA...GGCAGCAGCAATTCAGTTCGACCGTACG 317
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
72 ValMetGluIleGluAspGlySerAspThrSerAlaThrSerGlu.... 86
318 GCGCCGAGCGCTGGCAAACTTAAGCGCGGANGANTNNGNCAATCTGA 367
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
87 ProLysAla..... 89
368 TCCATTCGCGTTTGGACTCGCTGCGTANCCGCTTCGACCAATC 417
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
90 .....GluThrLysSerGluAlaLysVal 97
418 CCTGCGCTGATGCCGAGCGCTTGCATCTTCGTCATGCGATGAGAC 467
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
98 GluValValAlaGluAsn.....AlaSerValAlaGlyAl 109
468 CAATCCGCTNCGGAGACCGCTGTGTGATCAAGAACCCGANG 517
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
109 aThrProValSerAsnAsp.....ValIleValArgLysGlnThr 124
518 ATTTCAGACGANGTGTGTGATGACCGCTTGACCGACGCAATC 567
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
124 hrValAsnLysSerSerThrIleLysAlaThrProLeuAlaArgLysVal 140
568 CATGTGTGATGACGCTGCGCGACGATG..... 597
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
141 .....AlaAlaAspLeuAsnIleAspLeuSerLeuValThrPr 153
598 .....CCGTCTGAAT.....GCTGCCAATGCAACACATGAA 633
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
153 oThrGlyProAsnGlnArgIleLeuValAlaAspIleLysAsnHisGln 170
634 .....TTGGCGCGCCCGCAGCCGCC 654
170 laSerSerThrGlnLeuAlaSerGlnProIleSerGlnProAlaProThr 186
655 GGTTCGATGAGCAGCGAC.....ATTCAATTCATGGA 686
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
187 ProSerProSerAlaHisGlnThrIleAlaProThrIleLysValAlaG 203
687 GCCCGTGGT..... 696
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
203 upProSerAlaProLeuSerThrAspGluValProMetAsnGlyValArg 220
697 ..GCAACAAACCGTTTGACCATTAATCAAGATGAATTCGATC 744
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
220 ySalatThrValLysAlaMetThrLysSerHisThrGluIleAlaAla 236
745 GGACGTTTGTTCACAGCGCGCTGACAGCAGCGCGCTGATTCGTT 794
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
237 .....ThrGlyMetLysAsnThrAsp..... 243
795 GGGTGTTCATCACTCAACAAACCGCTTCGTCACGTTTG... 840
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
244 .....IleThrGlnThrHisLysMetArgThrGlnLeuLys 256
841 .....GGTGCAGAGTATCGCAAT..... 861
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
256 spHisAlaAlaAlaSerGlyLysLeuThrGlyLeuAlaPheIle 272
862 .....ACTGGGGGGAATTCGTGACGACAGCAACCGCGTATTCGCG 905
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
273 LysAlaValAlaLysSerLeuArgAspMetProAsnIleAsnValArg 289
906 TTCGTTTGAACGCGCGAT..... 927
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
289 yAspPheAlaAsnAsnLysIleGlnPheMetHisAsnIleAsnIleGly 306
928 .....ACACAAGCGCG 939
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
306 leaIaValaAspThrProAsnGlyLeuMetValProValIleLysGlyAla 322

```

```

940 CAGCATTTATTTGGAGCTACCAATCAATTCGTTATCGAAGAC 989
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
323 ...AspHisLeuSerValPheGlnIleAlaIleLysIleSerGlu.... 336
990 CCGCAGCAAAAGACTGTGCGTGGTGGCGCGGACCGCAATACT 1039
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
337 .....LeuAlaAsnLysAlaLysAspGly 345
1040 CCATCAGCGCTACG.....ACCGTCGGCAATTC 1068
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
345 yLeuThrArgAlaGluMetThrGluAlaThrPheThrValSerAsn 361
1069 CTGAACAAACAACTTCATCAATCAGACAGCGGTACAGCGCG 1118
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
362 GlySerValGlyLeuAspTyrAlaThrProIleIleAsnSerProGlu 378
1119 CGCCATGTCGCGATGTGATTCAGCGCG..... 1149
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
378 rAlaIleLeuGlyValGlyThrMetSerGlnThrProLeuTyrIleAsn 395
1150 .....GTATGCCGCTAGACATC 1167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
395 yGluLeuGlnLysArgPheIleMetProLeuSerMet 407

seq_name: swissprot_40:PMAL_DUNAC
seq_documentation_block:
ID PMAL_DUNAC STANDARD; PRT; 1103 AA.
AC P54210;
DR 01-OCT-1996 (Rel. 34, Created)
DR 01-OCT-1996 (Rel. 34, Last sequence update)
DR 16-OCT-2001 (Rel. 40, Last annotation update)
DE Plasma membrane ATPase (EC 3.6.3.6) (Proton pump).
GN DHAL.
OS Dunaliella acidophila.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Dunaliellaceae; Dunaliella.
OX NCBI_TaxID=38272;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SAG 19.85;
RA MEDLINE=97127854; PubMed=8972605;
RT Weiss M., Pick U.;
RT "Primary structure and effect of pH on the expression of the plasma
RT membrane H(+)-ATPase from Dunaliella acidophila and Dunaliella
RT salina."
RL Plant Physiol. 112:1693-1702(1996).
CC -!- FUNCTION: THE PLASMA MEMBRANE ATPASE OF PLANTS AND FUNGI IS A
CC HYDROGEN ION PUMP. THE PROTON GRADIENT IT GENERATES DRIVES THE
CC ACTIVE TRANSPORT OF NUTRIENTS BY H+-SYMPORT. THE RESULTING
CC EXTERNAL ACIDIFICATION AND/OR INTERNAL ALKALIZATION MAY MEDIANTE
CC GROWTH RESPONSES.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(IN) = ADP + PHOSPHATE +
CC H(+)(OUT).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC (E1-E2 ATPASES). SUBFAMILY IT1A.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: U54690; AAB49042.1; -.
CC InterPro: IPR004014; Cation_ATPase.
CC InterPro: IPR001757; E1-E2_ATPase.
CC InterPro: IPR000695; HATPase.
CC InterPro: IPR001454; Hydrolase.
CC Pfam: PF00690; Cation_ATPase_N; 1.

```

DR Pfam; PF00122; E1-E2_ATPase; 1.
 DR Pfam; PF00702; Hydrolyase; 1.
 DR PRINTS; PR00119; CATAPASE.
 DR PRINTS; PR00120; HATPASE.
 DR PROSITE; PS00154; ATPASE_E1_E2; 1.
 KW Hydrolyase; Hydrogen ion transport; Transmembrane; Phosphorylation;
 KW Magnesium; ATP-binding.
 FT TRANSMEM 80 100 POTENTIAL.
 FT TRANSMEM 101 121 POTENTIAL.
 FT TRANSMEM 154 174 POTENTIAL.
 FT TRANSMEM 232 252 POTENTIAL.
 FT TRANSMEM 266 286 POTENTIAL.
 FT TRANSMEM 306 326 POTENTIAL.
 FT TRANSMEM 643 663 POTENTIAL.
 FT TRANSMEM 690 710 POTENTIAL.
 FT TRANSMEM 734 754 POTENTIAL.
 FT TRANSMEM 845 865 POTENTIAL.
 FT TRANSMEM 881 901 POTENTIAL.
 FT TRANSMEM 944 964 POTENTIAL.
 FT MOD_RES 358 358 PHOSPHORYLATION (BY SIMILARITY).
 FT METAL 616 616 MAGNESIUM (BY SIMILARITY).
 FT METAL 620 620 MAGNESIUM (BY SIMILARITY).
 SQ SEQUENCE 1103 AA; 119790 MW; 6F8BC2370D03DDC8 CRC64;

alignment_scores:

Quality: 100.00 Length: 454
 Ratio: 0.467 Gaps: 25
 Percent Similarity: 47.137 Percent Identity: 20.925

alignment_block:

US-09-303-518d-127 x PMAL_DUNAC ..

Align seg 1/1 to: PMAL_DUNAC from: 1 to: 1103

```

28 CTGCCCATCGCGGCGAGACGCAATCATTTTGACGGGCGGCTC.. 75
   ||||| |||  ::  ||||| |||||
165 LeuProGluHnIslgYAlaAsp..  ..  TyrGluThrProValGI 198
   |||  ::||| ||||| ||||| |||||
76  ATTACCGAAGTCGCGTTCGTTGCGAAGAAATGCGGTATGCGCCCT 124
   |||  ::||| ||||| ||||| |||||
198 ntleAspGlnAlaIalaLeuThrGlyGluSerLeuProAlaLysPheT 215
   |||  ::||| ||||| ||||| |||||
125 NGATGAAGTCAAG...GAAGGCGATGCGGTCAAAAAGGCCAAGTGTG 171
   ::|||  ::||| ||||| ||||| |||||
215 hrcGlyAsnValAlaPheSerGlySerThrValLysGlnGlyL.. 229
   |||  ::||| ||||| ||||| |||||
172 TTTGAAGCAAAAAGNATCGCGGCGGTTCACCGCCGCGCTTCA.. 219
   |||  ::||| ||||| ||||| |||||
230  ..  ArgHisAlaValAlaValAlaValAlaValAlaValAlaVal 241
   |||  ::||| ||||| ||||| |||||
220  ..  GCGAAATCGCGGCGCATCGCGGCGAAAGCGCTA... 258
   ||||| ||||| ||||| ||||| |||||
241 rPhePheGlyAlaAlaIalaLeuIleSerGlyThrHisAsnValAla 258
   |||  ::||| ||||| ||||| |||||
259  ..CTTCAGTCGCGTGCATTCGCGGTGAAGC..... 288
   |||  ::||| ||||| ||||| |||||
258 snlleGlnValMetAsnArgIleGlyGlyLeuGlyLeuIleThrIle 274
   |||  ::||| ||||| ||||| |||||
289  ..  AACGACAAATCGATTGCGAAGCTACGCGGC 320
   |||  ::||| ||||| ||||| |||||
275 GlyValITPrValValIleGlnValProValGlnPheAlaHisTYFLYShI 291
   |||  ::||| ||||| ||||| |||||
321 CGAAGCGTTGGCAACTTAAGCGCGGAGGAAATNGNNGCAATGTGATCC 370
   ||||| ||||| ||||| ||||| |||||
291 sSerGlyValAlaGlyLysGlnGlyCysProThrLeuLysMetLeu.. 307
   |||  ::||| ||||| ||||| |||||
371 AATCGGTTTGAGCTGCGCTGCTANCGCTCGTTAGCAAAATCCCT 420
   |||  ::||| ||||| ||||| |||||
307  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  .. 307
421 GCGGTGATGCGGAGCGGTCGCGCATCTTGCAATGCGATGAGACCA 470
   ||||| ||||| ||||| ||||| |||||

```

```

308  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  .. 314
471 TCAGCTGCGGCGAGACCGCTGCTGTGATCAAAAGACCGGANGATT 520
   ||||| ||||| ||||| ||||| |||||
315  ..ProIleAlaMetProThrValLeuSerValIThrLeuAlaLeuGlyAlaT 331
   |||  ::||| ||||| ||||| |||||
521 TCAGA.....CGANGTNGCTGATATGAGCGGTTGACCGAGCTAAA 564
   |||  ::||| ||||| ||||| |||||
331 YrLysLeuAlaArgGlnGlyAlaIleValIThrArgMetSerAla..... 345
   |||  ::||| ||||| ||||| |||||
565 ATCCATGTGTGAAGCAGCTGGCGGAGACGTCGCTGAAATCTGCG 614
   |||  ::||| ||||| ||||| |||||
346  ....ValGlnIleMetAlaGlyLeuAspValLeuGlySerAspLysTh 360
   |||  ::||| ||||| ||||| |||||
615 CAACATGAAACCAATGAAATTCGCGCGGACATCCGCGGTTGAGTG 664
   |||  ::||| ||||| ||||| |||||
360 rGlyIThrLeuIThrLeuAsnLysLeuSerIleAspProSerAsn..... 374
   |||  ::||| ||||| ||||| |||||
665 GCAAGCAGATTCATTGATGAGCGCGTGGTCAAAAACCAACCGTTGG 714
   |||  ::||| ||||| ||||| |||||
375  ....ValPheProValGly..... 379
   |||  ::||| ||||| ||||| |||||
715 ACCATCATATATGAAATGTAATTCGATCGGACGCTTGTTCACACAG 764
   ||||| ||||| ||||| ||||| |||||
380 ThrMetAspIleProGlnValMetLysPheGlyAlaLeuSerAla.... 394
   |||  ::||| ||||| ||||| |||||
765 CCGTGTGACACCGGAGCGGTGATT.....GCTTGGGTGCTTCAAG 808
   ||||| ||||| ||||| ||||| |||||
395  ..AsnIleIleIThrGlnGluProIleAspMetValLeuITrGlnSerTYrP 411
   |||  ::||| ||||| ||||| |||||
809 TCAACAAACCAACCGCTCTTGGCT..... 831
   |||  ::||| ||||| ||||| |||||
411 rGlnITrGlnLysLeuLysSerGlnITrLysHisIThrLysTYrPhePro 427
   |||  ::||| ||||| ||||| |||||
832  ....  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  .. 866
428 PheAsnProAsnAspLysIleIThrIleAlaIThrValLeuGlnIleAlaTh 444
   ||||| ||||| ||||| ||||| |||||
867 GGGCGAATTGTTGACGCGAGACACACCGCGATTTCCGCTGCG..... 909
   ||||| ||||| ||||| ||||| |||||
444 rGlyITrGlnValPhe.....ArgValLeuLysGlySerITrGlnITr 457
   |||  ::||| ||||| ||||| |||||
910  ....  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  .. 930
457 alValleuAlaLysAlaITrPasnAlaGlnAlaLeuAspGlyProValAsn 473
   ||||| ||||| ||||| ||||| |||||
931 CAAGGCGGCGAGATTAATGGAGCGGTACCAACATCAGATTCGCTG.. 978
   ||||| ||||| ||||| ||||| |||||
474 GlnLysITrLeuSerITrAlaGlyITrGlyITrPheAlaGlySerLeuITrAla 490
   |||  ::||| ||||| ||||| |||||
979  ..ATCGAAGAAGCGCGAGCAAAAGCTGTCGCGTGG.....G 1015
   ||||| ||||| ||||| ||||| |||||
490 amETalagLugLysAspGlyLysAspGlyITrLysITrGlnMetLeuAlaVal 507
   |||  ::||| ||||| ||||| |||||
1016 TTGCGCGG.....CAGCGGAGCAAAATCTCCATCAGCGGTACGACCCCTC 1059
   |||  ::||| ||||| ||||| |||||
507 alLeuProMetPheAspProITrAlaGlnAsp...ThrLysGlnITrIle 522
   |||  ::||| ||||| ||||| |||||
1060 GCGCATTCCTGAAAAAACAATC...TTCAGTTCAGACAGCGCTCAA 1106
   |||  ::||| ||||| ||||| |||||
523 GlnITrGlySerMetLysGlnGlyITrAlaValLysMetValIThr..... 536
   ||||| ||||| ||||| ||||| |||||
1107 CCGTGGCGAGCGCGGCGATGCTGCGATTCGATTCAGACGCGGTA...A 1153
   ||||| ||||| ||||| ||||| |||||
537  ....GlyAspHisLeuITrIleGlyLysGlnITrAlaLysMetLeuGln 552
   |||  ::||| ||||| ||||| |||||
1154 TGCGGCTAGACATCTGCTGACCGCTGTTTGCGGATTAATCGCGCG 1203
   |||  ::||| ||||| ||||| |||||
552 etGlyIThrGlnMetLysITrProSerGlnValLeuITrLysAlaArgAsnGly 568
   |||  ::||| ||||| ||||| |||||
1204 GATACGAGACAG 1215
   |||  ::||| ||||| ||||| |||||
569 AspValGlnAla 572
   |||  ::||| ||||| ||||| |||||

```

```

seq_name: SwissProt_40:YGI_F_YEAST
seq_documentation_block:
ID YGI_F_YEAST STANDARD; PRT; 551 AA.
AC P53214;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE Hypothetical 57.5 kDa protein in YMA7-RPS25A intergenic region.
GN YGI023W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=8288C;
RX MEDLINE=97435481; PubMed=9290212;
RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
   chromosome VII."
RL Yeast 13:1077-1090(1997).
CC -1- SIMILARITY: TO YEAST MID2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z72807; CA97006.1; -.
DR SGD: S0003255; MTU1.
KW Hypothetical protein; Transmembrane.
FT DOMAIN 104 107 POLY-SER.
FT DOMAIN 111 120 POLY-SER.
FT DOMAIN 124 138 POLY-SER.
FT DOMAIN 142 146 POLY-SER.
FT DOMAIN 182 188 POLY-SER.
FT DOMAIN 215 218 POLY-SER.
FT DOMAIN 222 232 POLY-SER.
FT DOMAIN 236 264 POLY-SER.
FT DOMAIN 269 275 POLY-SER.
FT DOMAIN 278 281 POLY-SER.
FT DOMAIN 284 291 POLY-SER.
FT TRANSMEM 362 382 POTENTIAL.
FT DOMAIN 468 474 POLY-ASP.
SQ SEQUENCE 551 AA; 57527 MW; 86D943341B319951 CRC64;

alignment_scores:
Quality: 99.50 Length: 199
Ratio: 0.865 Gaps: 7
Percent Similarity: 57.789 Percent Identity: 24.623

alignment_block:
US-09-303-518D-127/rev x YGI_F_YEAST ..
Align seg 1/1 to: YGI_F_YEAST from: 1 to: 551
635 AATTCATGTTGATGATGTCGACGATTTTCAGACGCGACGCTGCGGCC 586
      ::||| ::||| ::||| ::|||
128 SerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 144
      ::||| ::||| ::||| ::|||
585 AGTCGCTTACACACATGATTTACGCTCGCTC..... 552
      ::||| ::||| ::||| ::|||
144 rSerSerLeuProThrPheThrValAlaSerThrSerSerThrValAla 161
      ::||| ::||| ::||| ::|||
551 ..AAAGCGCTCAATACACAGACGATTCGTCGAATCTGCGCGCTTCT 504
      ::||| ::||| ::||| ::|||
161 erSerThrLeuSerThrSerSerSerLeuValIleSerThrSerSer 177

```

```

503 TTGATCAACACACAGGCTGCGGACGAGGATGTCATTCGCAAT 454
      ::||| ::||| ::||| ::|||
178 ThrPheThrPheSerSerSerSerSerSerSerSerSerSer 194
      ::||| ::||| ::||| ::|||
453 GACGAGATGCGGACGAGCTCGGATCAGACGCGGATTTGCTGACG 404
      ::||| ::||| ::||| ::|||
194 SerThr.....SerValSerThr..... 200
      ::||| ::||| ::||| ::|||
403 GACGAGTACCGACGACGATCCACAAACGGATGATGACATTCGACNN 354
      ::||| ::||| ::||| ::|||
201 .....SerSerValThrValProSer.....SerSer 209
      ::||| ::||| ::||| ::|||
353 ANTTCNTGCGCGCTTAAGTTGCGCAACGCTTGGGGCGGAGCTGCA 304
      ::||| ::||| ::||| ::|||
210 ThrSerSerPro.....ProSerSerSerGluLeuThrSer 223
      ::||| ::||| ::||| ::|||
303 CTGATTTGCTGCTTGGCTTCAACGCAATCAGACGACGATGATGAC 254
      ::||| ::||| ::||| ::|||
223 rSerThrSerSerSerSerSerSerSerThrLeuPheSerThrSer 240
      ::||| ::||| ::||| ::|||
253 GCTTTTGGCCGCGATGATGCGGCGCATTTGCTTGAACGCGCGGTA 204
      ::||| ::||| ::||| ::|||
240 erPheSer.....SerSerSerSerSerSerSerSerSer 252
      ::||| ::||| ::||| ::|||
203 AACACACAGCGCGGATTCCTTTGCTTCAACGACGCTTGGCTTTT 154
      ::||| ::||| ::||| ::|||
253 SerSerSerSerSerSerSerSerSerSerSerThrLeuSer 269
      ::||| ::||| ::||| ::|||
153 GACGCGATCGCTTCTTCACTTCATCAGGCGCATACCGCGCATAT 105
      ::||| ::||| ::||| ::|||
269 rThrSerSerSerSerSerSerSerSerSerSerSerThrPhe 286
      ::||| ::||| ::||| ::|||
104 ....TCTTGGCCACACGACGATTCGTCATGACGCGCGGCTGA 63
      ::||| ::||| ::||| ::|||
286 erSerSerSerSerSerSerSerSerSerSerSerThrSer 301
      ::||| ::||| ::||| ::|||

seq_name: SwissProt_40:HFC1_MESAU
seq_documentation_block:
ID HFC1_MESAU STANDARD; PRT; 2090 AA.
AC P51611;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Host cell factor C1 (HCF) (VP16 accessory protein) (HFC1) (VCAF)
DE (CFF).
GN HFC1.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN 1;
RP SEQUENCE FROM N.A.
RA Goto H., Nishitani H., Umene K.I., Nakabeppu Y., Nishimoto T.;
RT Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: UPON LYTIC INFECTION OF PERMISSIVE CELLS, THE HSV
CC TRANSCRIPTIVATOR PROTEIN VP16 ASSOCIATES WITH HCF. BINDING TO HCF
CC ACTIVATES VP16 FOR ASSOCIATION WITH THE OCTAMER MOTIF-BINDING
CC PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX RESPONSIBLE FOR
CC ACTIVATING TRANSCRIPTION OF THE HSV IMMEDIATE EARLY GENES (BY
CC SIMILARITY).
CC -1- SUBUNIT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM
CC 110 TO 150 KDA AND A MINOR 300 KDA POLYPEPTIDE. THE MAJORITY OF N-
CC AND C-TERMINAL CLEAVAGE PRODUCTS REMAIN TIGHTLY, ALBEIT
CC NONCOVALENTLY, ASSOCIATED (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DOMAIN: THE HCF REPEAT IS A HIGHLY SPECIFIC PROTEOLYTIC CLEAVAGE
CC SIGNAL (BY SIMILARITY).
CC -1- PTM: EXISTS AS A FAMILY OF POLYPEPTIDES THAT ARISE THROUGH
CC CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS
CC AT A DEFINED SITE, PCPE/THET, WITHIN THE HCF REPEAT (BY

```


CC SIMILARITY).

CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; D45419; BAA08258.1; .

DR HSSP; P02751; IPNA.

DR InterPro; IPR003961; FN.III.

DR InterPro; IPR001798; Kelch.

DR Pfam; PF000041; fn3; 2.

DR Pfam; PF01344; Kelch; 4.

DR SMART; SM00060; FN3; 2.

KW Nuclear protein; Repeat.

FT REPEAT 44 89

FT REPEAT 93 140

FT REPEAT 148 194

FT REPEAT 217 265

FT REPEAT 266 313

FT DOMAIN 1010 1448

FT REPEAT 1010 1035

FT REPEAT 1072 1097

FT REPEAT 1101 1126

FT REPEAT 1157 1182

FT REPEAT 1295 1320

FT REPEAT 1323 1348

FT REPEAT 1358 1383

FT REPEAT 1423 1448

SO SEQUENCE 2090 AA; 214942 MW; E495B8B1F2385E17 CAC64;

alignment_scores:

Quality: 99.00 Length: 507

Ratio: 0.446 Gaps: 23

Percent Similarity: 43.787 Percent Identity: 20.118

alignment_block:

US-09-303-518D-127/rev x HFCL_MESAU

Align seg 1/1 to: HFCL_MESAU from: 1 to: 2090

1189 CCGCGAAAGCAGGAGTGTCTAGCGGATACGGCGTCTGTA 1140

|||||:|||||

320 ArgAlaArgAlaGly.....HisCysAlaValAlaIleAsnThrArgLe 334

1139 GTACCAATGCGACCATGCGCGGTGCGCGTGTGCGTGA 1090

|||||:|||||

334 uTyr.....IleTyrSerGlyArgAspGlyTyrArgGlyAlaTprA 348

1089 GTTGAAGAGTTGTTTTCAGG..... 1068

||:|||||

348 snAsnGlnValCysCysLysAspLeuTprTyrLeuGlnThrGlnLysPro 364

1068 1068

365 ProProProAlaArgValGlnLeuValArgAlaAsnThrAsnSerLeuGln 381

1067AAATGGCGGAGGTGTGACGCGTGTGATGAGATTGTTG..... 1032

|||||:|||||

381 uValSerTprGlyAlaValAlaThrAlaAspSerTyrLeuLeuGlnLeuG 398

1031TCGGGCTGGCGGCAACCGCAACCGCAACGAC 1002

|||||:|||||

398 lntStyAspLleProAlaThrAlaAlaThrAlaThrSerProThr... 413

1001 TCTTTCGTCGGCGCTTCTTCGATTAACGGAATCTGATGTGTGCTACGCTGC 952

|||||:|||||

414ProAsnProValProSerValProAla.....AsnPr 424

951 CAAATATCGTGGCGCGCTTGTGTAATCGCGCG..... 918

|||||:|||||

424 oProLysSerProAlaProAlaAlaAlaProAlaValGlnProLeuT 441

917TTCAATACGGAACCGGAATACGCGGTTGTCTGCGTCA 879

|||||:|||||

441 hrgInValGlyIleThrLeuValProGlnAlaAlaAlaProProSer 457

878 ACCAATTCG.....CCCCGAGTAATTTGGGATTACTTTCGCACC 841

|||||:|||||

458 ThrThrThrThrIleGlnValLeuProThrValProGlySerIleSerVal 474

840 CAAACCGTACGCAAGAGCGGTGT.....TTGTTGACTTTGGAAC 800

|||||:|||||

474 lProThrAlaAlaArgAlaGlnGlyValProAlaValLeuLysValThrG 491

799 CACCCAAAGCAATACGCGCTCG.....GTGTTCAAGCGCGCTGTT... 759

|||||:|||||

491 lProGlnAlaThrThrGlyThrProLeuValThrMetArgProAlaGly 507

758 ...GCAACAAAGCTCGGATGCGAATTTACATCTGTAATTAATGCTGCA 712

|||||:|||||

508 GlnAlaGlyLysAlaProValThrValThrSer.....LeuProAlaSe 522

711 AACGTTTGTGTTGACACCGGCGTCAATGAATGAATGCGTGCAC 662

|||||:|||||

522 rValArgMetValValProThrGlnSerAlaGlnGlyThrValIleGlyS 539

661 TCAACCGCGCGGATGCGG..... 642

|||||:|||||

539 eAsnProGlnMetSerGlyMetAlaLeuValAlaAlaAlaAlaAla 555

641CCGCGAATTCATGCTTTCGATGTTGCA..... 612

556 ThrGlnLysIleProProSerSerAlaProThrValLeuSerValProAl 572

611GCATTTCGACGACGACGCTGCGC 587

572 aglyThrThrIleValLysThrValAlaValThrProGlyThrThrThr 589

586 CAGCTGCTTACACACATGATTTTACGCTGCTGCAACGCGTCAATACC 537

|||||:|||||

589 euProAlaThrValLysValAlaLysSerSerProValMetValSerAsnPro 605

536 AGCANACNTGCTGTAATGTCGCGCGTCTTGTGATGCACACACAG 487

|||||:|||||

606 AlaThrArgMetLeuLysThrAlaAlaAlaGlnValGlyThrSerValSe 622

486 GTCTGCCGN.....AGCG 473

|||||:|||||

622 rSerAlaAlaAsnThrSerThrArgProIleIleThrValIleLysSerG 639

472 GATTGGTGCATCGCA.....TTGACGAAGATGGCGAAC 438

|||||:|||||

639 lYthrValIthrValAlaGlnGlnAlaGlnValAlaValIthrValIgly 655

437 GCGTCGCGATCGACGCGGAGGATTTGCTGACGAGCGAGTACGACGCG 388

|||||:|||||

656 GlyValThrLysThrIleThrLeuVal.....LysSerTrp 667

387 AGTCACAAACCGGATGTGATCAGATTGCNNNN..... 354

|||||:|||||

667 oLleSerValProGlyLysSerAlaLeuLleSerAsnLeuGlyLysValM 684

353ANTTCNTGCCCGCTTAAGTTTCCCAACGCTTGGCGCGC 315

|||||:|||||

684 etSerValValGlnThrLysProValGlnThrSerAlaValIthrGlyGln 700

314 TAGCGTTGCAACTCGATTTGCTGTTG..... 288

|||||:|||||

701 AlaserThrGlyProValIthrGlnIleIleGlnThrLysGlyProLeuPr 717